Elucidation of the sequence selective binding mode of the DNA minor groove binders SJG-136 and adozelesin, by high field 1H NMR and restrained molecular dynamics

Hopton, Suzanne

Award date:
2007

Awarding institution:
University of Bath

Elucidation of the Sequence Selective Binding Mode of the DNA Minor Groove Binders SJG-136 and Adozelesin, by High Field ^{1}H NMR and Restrained Molecular Dynamics.

Submitted by
Suzanne Hopton
for the degree of PhD
of the University of Bath
2007

The research work carried out in this thesis has been carried out in the Department of Pharmacy and Pharmacology, under the supervision of Dr. Andrew S. Thompson.

S. Hopton

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

The binding of two covalent minor groove binding ligands to duplex DNA has been investigated using high field ^{1}H NMR and NMR/NOE refined molecular modeling techniques. Various duplex DNA strands were synthesised and then reacted with the covalent minor groove binding drugs SJG-136 and adozelesin. The drug-DNA adducts were analysed using 2D NMR techniques and the resulting data used to produce NOE distance refined molecular models.

SJG-136 is a synthetic dimeric covalent minor groove binder based on the PBD family of antitumour antibiotics. These ligands are isolated from various *Streptomyces* species and exhibit potent *in-vitro* and *in-vivo* activity. They are known to form inter-strand cross-links with duplex DNA by reaction with the *exo*-cyclic NH_2 group of guanine. Adozelesin is a covalent minor groove binding analogue of the CPI antitumour antibiotic (+)-CC-1065. The CPIs are known to alkylate duplex DNA by reaction with the N3 position of an adenine base, and owe their biological activity to an ability to block DNA replication.

Four novel ligand-DNA adducts have been produced and fully assigned using sequential assignment techniques. The NMR data collected has been used to confirm the sites of alkylation, orientation of the ligand residue and stereochemistry of the interactions. NOE distances are then used to produce accurate NMR refined molecular models of each adduct.

The 5'*d*(CICGATCICG)$_2$-SJG-136 adduct was found to bind covalently to *exo*-cyclic NH_2 groups on guanine bases of opposite DNA strands – forming an inter-strand cross-link separated by 4 base pairs. Stereochemistry was assigned as 'S' at both reaction sites and self complementarity and β-helical structure of the DNA duplex was maintained. Molecular models show minimal distortion throughout the duplex, with only a small localised area of distortion in the central base pairs as a result of drug binding.

The 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct has been successfully produced, with alkylation sites confirmed as the *exo*-cyclic NH_2 groups of two guanine bases on the same DNA strand. This represents the first identification of an intra-strand

cross-linked PBD adduct, generating an exciting new possibility involving targeting of the human telomere repeat sequence using PBD type ligands. Once again stereochemistry at the reaction sites is confirmed to be 'S' in both cases and the drug is found, as expected, to associate more closely with the modified DNA strand than with its complementary sequence.

A 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct has been generated, and found to contain a mixture of adducts in an approximately 50/50 ratio. One adduct has retained Watson-Crick base pairing within the DNA duplex, while in the second the central AT step had adopted a Hoogsteen conformation. In both adducts a significant overlap and stacking of the benzofuran subunits is observed. The discovery of novel 'stacked' adducts for CPI ligands suggests an alternative model for minor groove drug interactions and similarities can be drawn between the adozelesin stacked molecules and analogous stacking in the non-covalent minor groove binding lexitropsin drugs.

ACKNOWLEDGEMENTS

There are many people I would like to thank for their help and support during my postgraduate studies. Firstly, I would like to thank Dr Andrew S. Thompson for providing me with the opportunity to study for a PhD at the University of Bath, Department of Pharmacy, and for the many hours of help and support that he has given. Department of Pharmacy, University of Bath for funding this post-graduate study. Dr Steven Black for his support and patience during my first year, Dr Timothy Woodman, for proof-reading this thesis and his many thoughtful comments and suggestions in addition to his assistance in the accumulation of the NMR experimental data.

In addition I would like to thank my family and friends for unfailing support during the past few years. In particular my parents for patience and support throughout this project, Mr Huw Roberts for all his encouragement and believing I could do it. Dr Adrian Neal for support during my time at Bath and his many helpful suggestions during the writing of this thesis.

CONTENTS

FIGURE CONTENTS

ABBREVIATIONS

1D	One-Dimensional
2D	Two-Dimensional
3'	Polarity directed to 3' phosphate terminal of DNA
5'	Polarity directed to 5' phosphate terminal of DNA
Å	Angstrom
A	Adenine
Ado	Adozelesin
'A' tract	Adenine rich oligonucleotide
C	Cytosine
COSY	Correlation Spectroscopy
CPG	Controlled Pore Glass
CPI	Cyclopropapyrroloindole
d	Deoxyribonucleotide
DNA	Deoxyribonucleic Acid
DQF	Double Quantum Filter
F_1	Frequency Domain 1
F_2	Frequency Domain 2
G	Guanine
HPLC	High Pressure Liquid Chromatography
Hz	Hertz
I	Inosine
Im	Imidazole
J	Coupling Constant
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
p	Poly *i.e.* poly adenine is p(A)
PBD	Pyrrolo[2,1-*c*][1,4]benzodiazepine
ppm	Parts Per Million
Pu	Purine
Py	Pyrimidine

RNA	Ribonucleic Acid
ROESY	Rotating Frame Overhauser Effect Spectroscopy
T	Thymine
t_1	Evolution Time
'T' tract	Thymine rich oligonucleotide

CHAPTER 1: INTRODUCTION

1.1 Cancer and Chemotherapy

1.1.1: What is Cancer?

Cancer is a broad term encompassing up to 200 neoplastic diseases, and is characterised by the presence of malignant cells. It is defined as "any of various malignant neoplasms characterized by the proliferation of anaplastic cells that tend to invade surrounding tissue and metastasize to distant body sites".[1]

In 2002 12.6 % of deaths in World Health Organisation member countries worldwide and 19.2 % of deaths in Europe were attributed to malignant neoplasms. As such cancer is the 2nd major cause of death in Europe and throughout the world.[2]

1.1.2: Causes of Cancer

Many causes of cancer are attributed to the formation of cancerous cells. These include the inheritance of faulty genes, exposure to radiation or carcinogenic chemicals, viruses, and diet.

Carcinogenic compounds in the environment include asbestos, compounds of beryllium, cadmium, chromium, nickel, lead and arsenic, aromatic amines, amino azo-dyes and alkylating agents.[3] Carcinogenic chemicals such as diethylamine are found in cigarette smoke and the mood altering tetrahydrocannabinol (THC) component of marijuana smoke has been found to increase DNA mutations by 300 %.[4] Chloroform, a by-product of the chlorination of water, is also a carcinogen[5,6] and drugs such as oral contraceptives have also been shown to increase the risk of breast cancer.[7]

Links between diet and cancer have been suggested numerous times through the media, with varying degrees of scientific evidence. Alcohol and foods contaminated with alfatoxins have been established as clear risk factors for some cancer sites. Excessive

alcohol consumption is known to cause inflammation and cirrhosis of the liver – leading to liver cancer.[8] A high intake of red meat has been suggested to lead to an increased risk of cancer of the stomach, colon and rectum[9] while a diet rich in fresh fruit and vegetables is thought to reduce the risk of cancer of the digestive tract and stomach. There are various suggestions that the replacement of saturated fats in the diet with unsaturated fats or olive oil also reduces the risk of cancer.

Exposure to ionising radiation has been related to the formation of some types of cancer and the exposure to UV light is well known to be responsible for a large number of skin cancers, including malignant melanomas and basal cell and squamous carcinomas.[8]

1.1.3: The Treatment of Cancer

Treatments for cancer include surgery, radiation therapy and cancer chemotherapy.

The surgical procedure involves the removal of the tumour and sometimes the complete affected organ. The surgery is followed by treatments such as chemotherapy and photoradiation therapy in order to increase chances of survival by shrinking or killing any cancer cells that were not able to be removed. Surgery is really only effective when a tumour is in its early stages or has not metastasised.

Radiation therapy involves the use of radioisotopes such as cobalt-60 and X-rays to destroy solid tumours. Tumours can often be destroyed with little damage to the surrounding tissue but the use of the treatment is limited due to the carcinogenic nature of the ionising radiation used.

1.1.4: Cancer Chemotherapy

Cancer chemotherapy is available in many forms, including anti-metastatic agents, antimitotic agents, antimetabolites, monoclonal antibody therapy, gene therapy, and DNA targeted agents such as DNA intercalators and alkylators.

Metastasis is comprised of growth and detachment from the primary site, local invasion and destruction of extra-cellular matrix, intravasation into blood vessels, survival in the circulation, extravasation from the vessels into the metastasis site and growth at the metastasis site.[11,12] Anti-metastatic agents are agents that inhibit enzymes involved in these processes in order to prevent tumours from metastasising.

Antimetabolites interfere with cell division and the growth of tumours. An example is Methotrexate – a dihydrofolate reductase (DHFR) inhibitor. DHFR is involved in the formation of tetrahydrofolic acid (FH4), which is required for the synthesis of DNA. Therefore this antimetabolite can prevent DNA synthesis and so lead to apoptosis.

Antimitotic agents are mainly naturally occurring compounds that prevent cell division and are therefore cytotoxic.

Monoclonal antibody therapy (MAB) is the use of highly specific antibodies to target cells. The treatment relies upon the stimulation of the patient's immune system to prevent tumour growth by blocking specific cell receptors. Radioactive and chemical toxins can also be attached to these antibodies. These are activated at the site of action and so affect only tumour cells.[10]

Gene therapy and antisense technology are promising cancer treatments that act at the molecular level. Antisense technology uses oligonucleotide drugs to bind to target RNA sequences, blocking the production of specific proteins. Once bound, the antisense agent either disables or induces the degradation of the target RNA.[13] Gene therapy involves the insertion of genes into cells and tissues of a cancer patient. It is possible that this could be useful in the reactivation of tumour suppressor genes in cancer cells where these genes have been deactivated.

1.1.5: DNA targeted agents

DNA targeted agents are drugs that target DNA in order to suppress or kill cancer cells. They include intercalating agents and alkylating agents. Both work by prevention of DNA transcription and replication. Examples of intercalating agents include

anthracycline antibiotics such as daunorubicin (daunomycin) and doxorubicin (adriamycin) (**Figure 1**).[14]

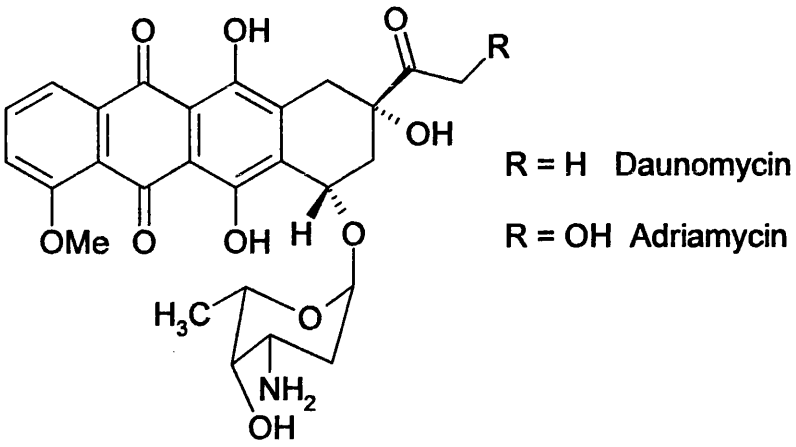


Figure 1: Chemical structure of two well-known anthracycline drugs

Alkylating agents are highly electrophilic molecules that react with nucleophilic centres of DNA bases – such as oxygen and nitrogen atoms. The first example of a cancer chemotherapy agent was the use of the nitrogen mustard alkylating agents in the 1940's.[15] An example of a drug with a similar mode of action is cis-platin (**Figure 2**).

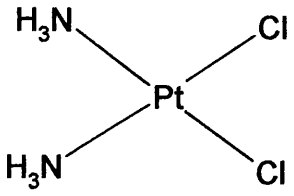


Figure 2: Chemical structure of cis-platin

Although this is not strictly an alkylating agent it cross-links DNA and inhibits DNA replication in a similar manner to the alkylating agents. Alkylating agents are the focus

of this project and so specific agents will be discussed in greater detail in subsequent sections.

1.2: Basic DNA Structure

DNA consists of a series of repeating units known as nucleotides. Each nucleotide consists of a DNA base (Adenine, Guanine, Cytosine or Thymine) attached to a 5-membered ribose sugar and joined to the next unit by a phosphate group. DNA bases are split into two categories – the purines, and the pyrimidines (**Figure 3**).

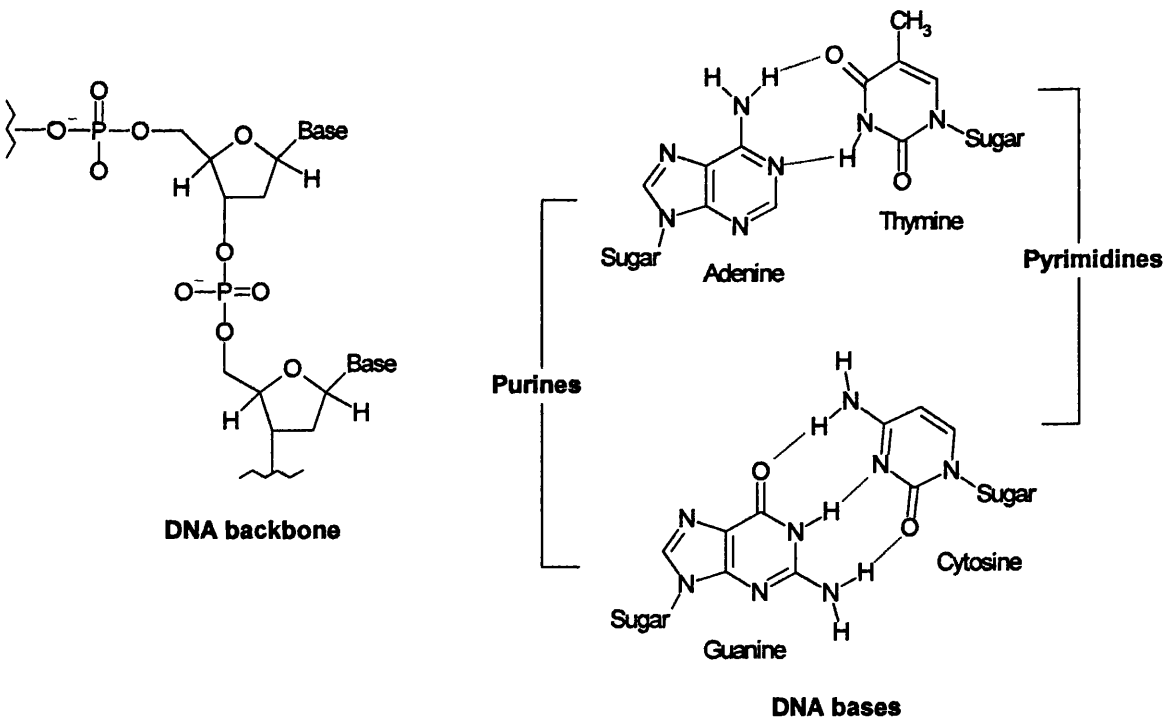


Figure 3: Basic Structure of DNA showing backbone, bases and unique hydrogen bonding

The bases bind specifically (G to C and A to T) by unique hydrogen bonds as shown in **Figure 3** to form the characteristic DNA double helix, a representation of which is shown in **Figure 4.**

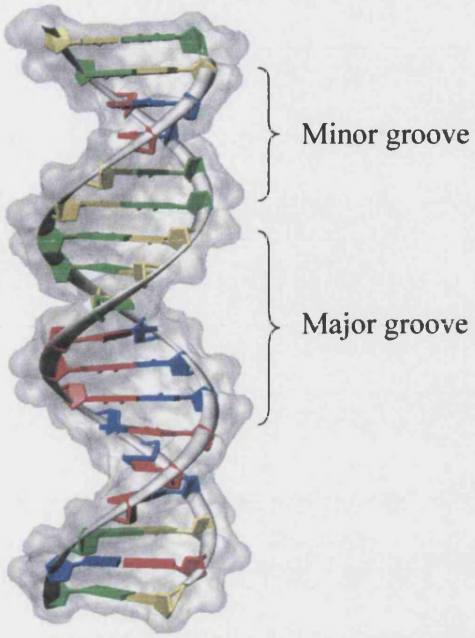


Figure 4: Representation of β-helical DNA, showing the major and minor grooves

1.3: DNA interactive ligands

There are three possible sites on B-form DNA to which a molecule can bind – the major groove, the minor groove or by intercalation between the bases. The major groove is deep and wide, lined with potential hydrogen bond acceptor atoms such as N7 of guanine and adenine, O4 of thymine and O6 of guanine. Hydrogen bond donor atoms also exist in the major groove in the form of the amino group at C6 of adenine and C4 of cytosine.[16] There are therefore a large number of potential hydrogen bond interactions. The major groove is also very accessible due to its width. As such this groove is often the site of choice for the interaction of many large proteins (such as transcription factors and enzymes). The depth of the major groove is conducive to small drugs binding but in practice the width of the site makes it unfavourable to many small ligands, particularly those with hydrophobic tendencies, which prefer a tighter fit into the DNA minor groove.

The minor groove of β-helical DNA is deep and narrow, so enabling small drugs to fit tightly within the hydrophobic environment. A large surface area is available, with opportunities for stabilising hydrogen bonds as well as many *Van der Waals* forces for small planar molecules. Hydrogen bond acceptor molecules of the minor groove consist of N3 of adenine and guanine and O2 of cytosine and thymine. A possible hydrogen bond donor site is the amino group of guanine.

The most important factor to the achievement of sequence selectivity in the minor groove is the topology. "A tracts" in the minor groove give rise to very narrow grooves for a close complementary fit with the ligand, while GC regions can accommodate bulkier molecules.

Interactions with DNA can be both covalent and non-covalent. Classes of compounds reacting with DNA can be split into four main types:

1. Intercalators/*bis*-intercalators such as Adriamycin and *bis*-acridines.
2. Minor groove binders such as cyclopropapyrroloindoles (CPIs), pyrrolo[2,1-*c*][1,4]benzodiazepines (PBDs) and lexitropsins.
3. Major groove binders including many proteins.
4. DNA cleavage agents - an example of which is bleomycin.

1.4: Minor Groove Binding Agents

Reactions within the minor groove take place primarily by a nucleophilic mechanism. Sites of nucleophilic reaction within the minor groove are shown in **Figure 5.**

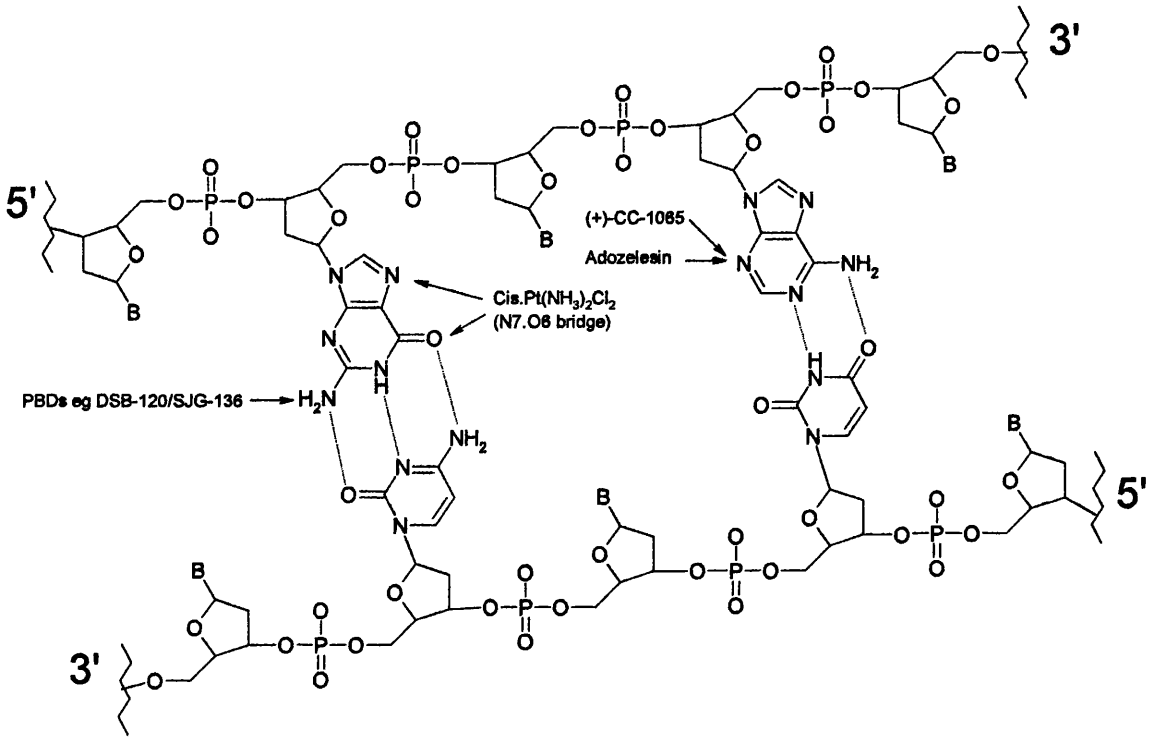


Figure 5: Schematic showing sites of nucleophilic reaction within the minor groove

Work in this project focuses upon the minor groove binders – namely PBDs and CPIs, with reference to netropsin and the lexitropsins. There follows a review of the history of these compounds and their binding to DNA.

1.5: Pyrrolo[2,1-*c*][1,4]benzodiazepines

1.5.1: History

The pyrrolo-[2,1-*c*][1,4]benzodiazepine (PBD) family are a series of potent antitumour antibiotics produced by various *Streptomyces* species. They were first discovered in 1965[17] when an antibiotic was isolated from *Streptomyces refuineus,* a thermophilic

actinomycete by Leimgruber *et al.*[18] The *actinomycete* had been shown two years earlier to produce a fermentation broth with antitumour activity.[19]

When characterised the antibiotic was found to consist of a tricyclic unit comprising a benzene ring, a pyrrole ring and a diazepine ring with a C ring substituent. The system bears a resemblance to anthranilic acid and so was named **anthramycin** and the new antibiotic group the **pyrrolobenzodiazepines** (PBDs). The structures of the naturally occurring PBDs subsequently discussed can be found in **Figure 6.**

Anthramycin was found to exhibit a remarkably potent antitumour activity[20] and the subsequent interest in this class of antibiotics led to the discovery of many more PBDs in the following years. In 1972 **tomamycin** and **sibiromycin** were isolated from *Streptomycin achromogenes* var. *tomamyceticus*[21] and *Streptosporangium sibirium*[22] respectively, although the suggested structure of sibiromycin was challenged and a new structure suggested in 1988.[23]

The next PBDs to be isolated were the **neothramycins** (A and B) in 1976.[24] 1980 saw the discovery of a similar PBD to Anthramycin, named **mazethramycin.**[25]

Throughout the 1980s several more PBD compounds were isolated and characterised. These included **prothracarcin,**[26] **DC-81,**[27] **chicamycin,**[28,29] **abbeymycin,**[30] **porothramycin**[31] and finally **DC-102 (sibanomycin).**[32]

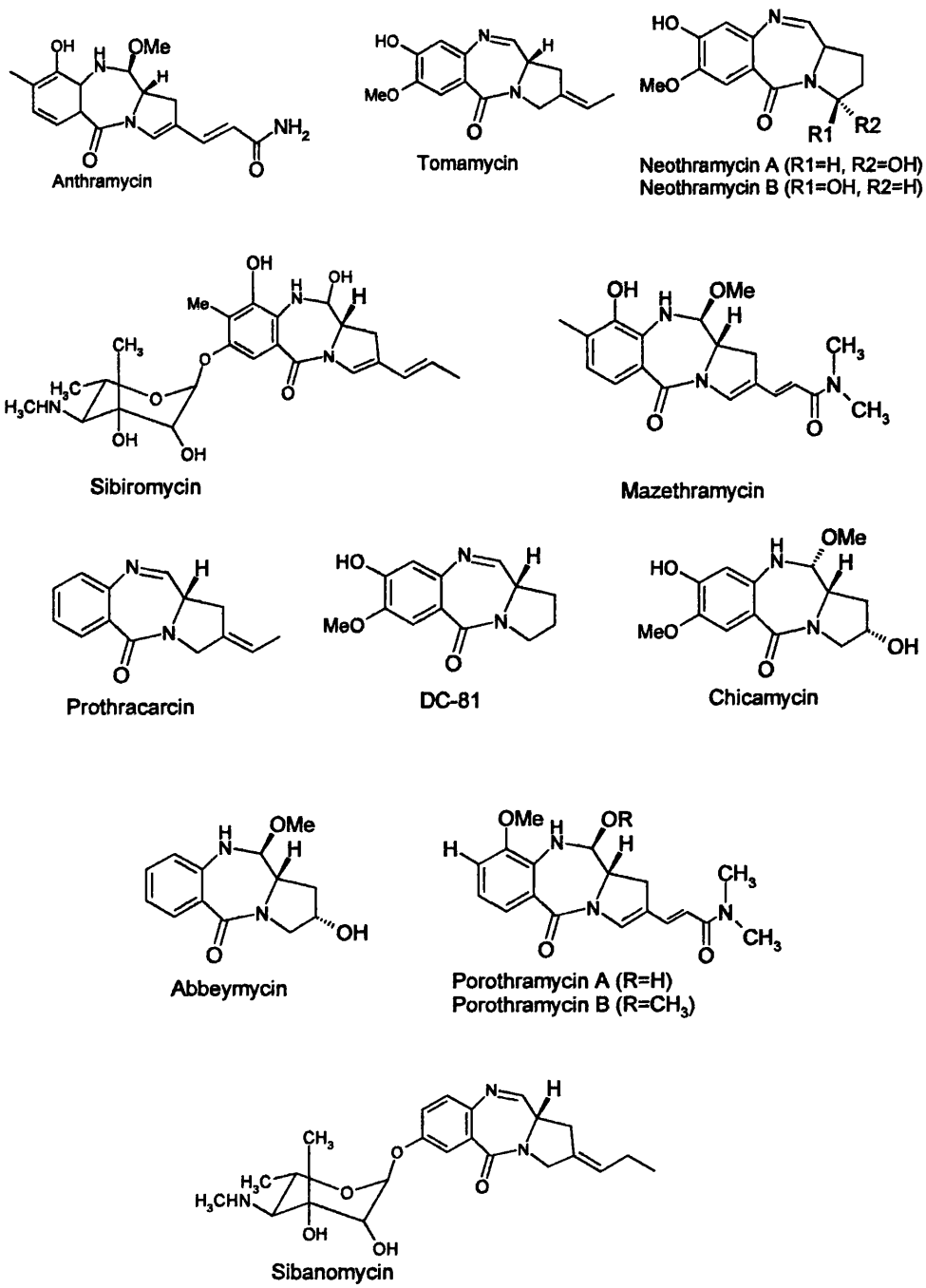


Figure 6: Chemical structures of the naturally occurring PBDs

1.5.2: Mode of Action

The PBDs are natural antitumour agents, with their biological effectiveness stemming from the inhibition of nucleic acid synthesis by covalent attachment to double stranded DNA. Binding occurs deep in the narrow minor groove covering a 3 base-pair region and requiring duplex DNA and the presence of guanine. Due to the shape of the PBDs they cause minimal disruption to the DNA helix. On reaction with PBDs the DNA helix becomes less flexible and appears to be stabilised, shown by an increased melting point on formation of adducts.[33-37]

PBDs are named and numbered according to the system designed and utilised by Leimgruber in 1965.[18] The numbering system for the PBD subunit is shown in **Figure 7.**

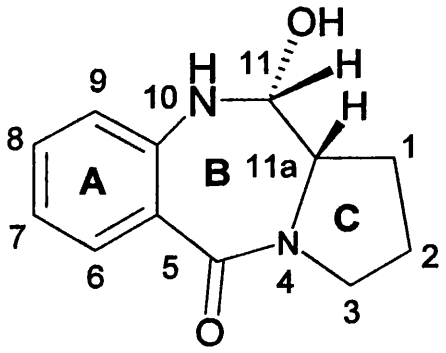


Figure 7: Numbering system for the PBD subunit showing the 'A', 'B' and 'C' rings[19]

Reaction with DNA is between the PBD-C11 position of the drug and the *exo*cyclic NH_2 group of guanine. The reaction site N10-C11 exists as either an imine, carbinolamine methyl ether or a carbinolamine form (**Figure 8**). The three forms are interchangeable and exist as a dynamic equilibrium, with the predominant species dependent on the compound structure, the synthetic work-up or the method of isolation. In a study involving anthramycin, tomamycin and the neothramycins[38] the imine functionality was shown to be the most reactive of these species.

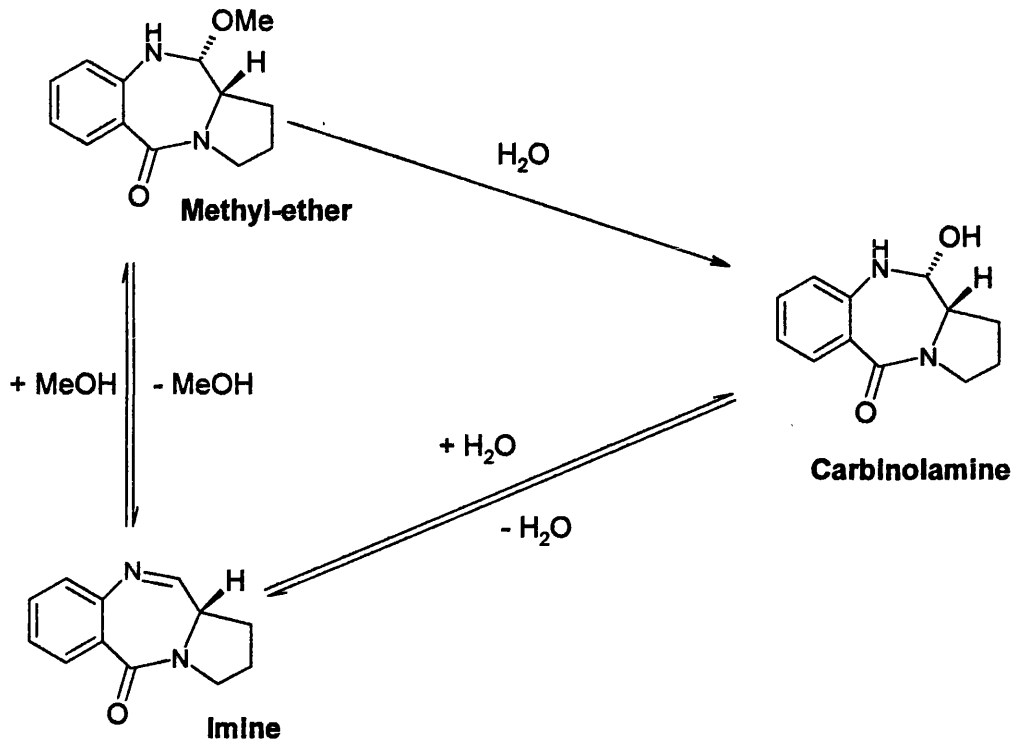


Figure 8 – Interchangeable forms of the PBD subunit

1.5.3: Binding of PBDs to DNA

All forms of the PBD subunit possess the ability to alkylate N2–guanine residues. Direct attack on the imine functionality as well as SN2-type mechanisms have been suggested for the reaction (**Figure 9**).[39]

The SN2 – type reaction consists of a nucleophilic attack by the NH_2 group of guanine on the carbinolamine or the methyl ether.[40] Alternatively there is a possibility of direct attack on the imine species.[41] It is however likely that the drug exists in the protonated form when present in the human body, so it is possible that the SN2 type reaction would be favoured under these circumstances.

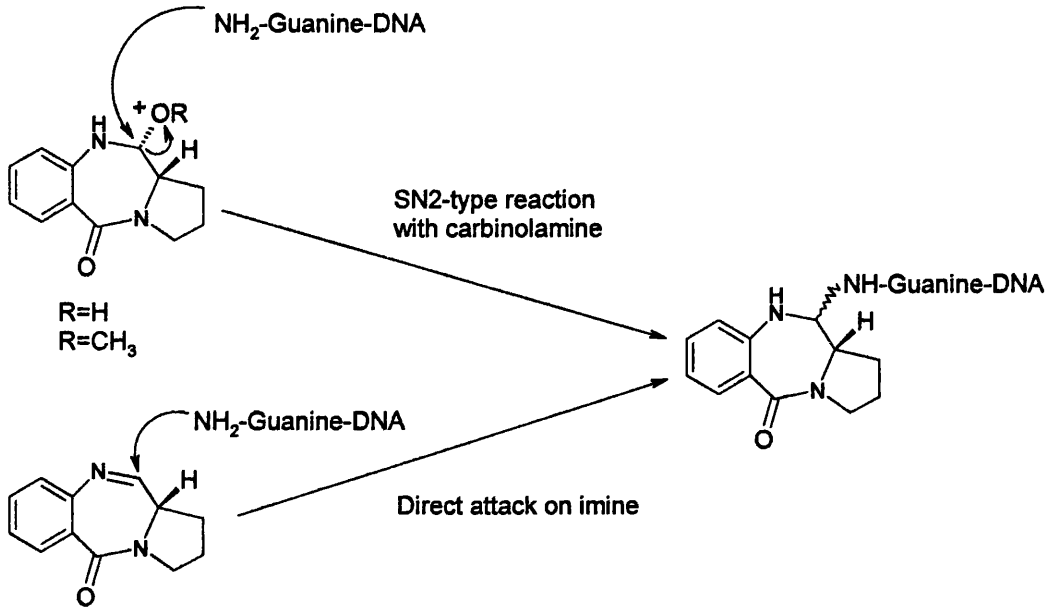


Figure 9: Two possible mechanisms for PBD adduct formation[40,41]

Molecular modeling of the PBD-DNA adducts predicts that the drug is completely submerged within the minor groove, following the right-handed contour of the B-DNA turn.[33]

1.5.4: Stereochemistry at the PBD chiral centres

In theory, the *exo*-cyclic amino group can attack the PBD-C11 position from either side. This results in possible 'R' or 'S' isomers of the drug-DNA adduct (**Figure 10**).

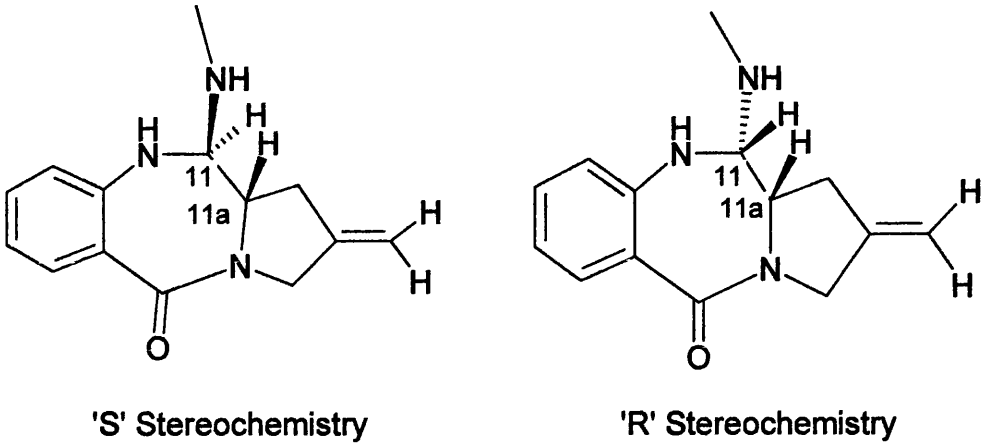


Figure 10: Representation of the PBD subunit showing 'R' and 'S' stereochemistry

However, molecular mechanics calculations have predicted that the lowest energy conformation is the PBD situated in the minor groove as an 11'S' isomer, with any side chain pointing toward the 5' end of the modified strand.[42, 43, 44] This has been confirmed by NMR experiments[45] that show the drug to be held strongly in one orientation.

A second chiral centre exists at PBD-C11a and all naturally occurring compounds have been found to display 'S' configuration at this point. This has the effect of giving the molecules a right hand twist from the 'C' to the 'A' ring that allows them to fit snugly into the minor groove of right-handed B-form DNA, with racemization at C11a significantly reducing the biological activity.[38,46] Previous experimentation has predicted that the 'S' configuration is probably stabilised by hydrogen bonding interactions between the drug and the oligodeoxynucleotide.[33]

1.5.5: Binding site preferences of the PBDs

Foot printing experiments have shown that drug adducts of PBD compounds span three base pairs, with the following sequence preference: 5'Pu-G-Pu > 5'Pu-G-Py or 5'Py-G-

Pu > 5'Py-G-Py.[38,46,47] The rank order of DNA reactivity for natural PBDs is as follows: sibiromycin > anthramycin > tomamycin > DC-81 > neothramycin.[48]

Although in general the PBDs react similarly with DNA some discrepancies have been noted. It was found that with anthramycin and tomamycin for instance, although the two drugs share one preferred binding site (-AGA-) they show different preferences when binding to other sites such as –GGC-, -GGG-, and –TGC-.[49] Many sites will bind to one drug but not to the other. DNA saturated with anthramycin has also been shown to still effectively bind tomamycin. This indicates that these drugs, although overlapping, also exhibit specific binding sites. Nearest-neighbour analysis of these adducts indicates that bases at the 3' and 5' sides of guanine are critical for the anthramycin and tomamycin binding process. It was found that G on the 3' side is least favourable for tomamycin but favourable for anthramycin. Therefore there are common factors governing the binding of these drugs, but also factors that will affect them differently.[49]

1.5.6: Anti-tumour Activity

Anti-tumour activity of the PBD family has been shown against transplanted tumours including *Ehrlich solid carcinoma, sarcoma 180, Human epidermoid carcinoma* and *Leukaemia L1210* cell lines.[50,51] In the clinical use of anthramycin and sibiromycin the most responsive tumours are gastrointestinal and breast neoplasms, lymphomas and sarcomas, with the least responsive ovarian, pancreatic carcinomas and malignant melanomas. No evidence for bone marrow depression or gastrointestinal toxicity has been found.[51] but members of the PBD group have displayed a dose-limiting cardiotoxicity similar to that observed with anthracyclines, adriamycin and daunomycin.[53] It is likely that the mechanism for cardiotoxicity is the same for both anthracyclines and the PBD family despite their structural differences, as protective therapy used in conjunction with anthracycline treatment is also effective against anthramycin-induced cardiotoxicity. Tautomerization or oxidation of the PBD subunit as shown in **Figure 11**[39] will lead to ortho-quinone products and it is these that are thought to be responsible for the observed cardiotoxicity. This is supported by structure-activity relationships, as anthramycin and sibiromycin, which both possess oxygen at the C9 position, are cardiotoxic.

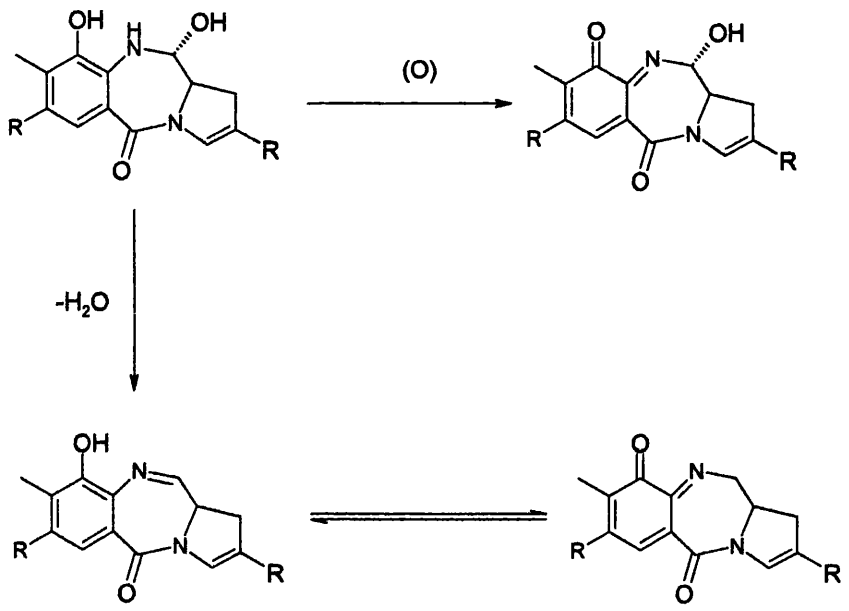


Figure 11: Postulated mechanism for quinoneimine formation[39]

1.5.7: Ring Substitution

Various studies have been carried out in order to determine the effects of substitution at the 'A' and 'C' rings of the PBD subunit. The importance of the 'C' ring for cytotoxicity was demonstrated as early as 1985[54] during an investigation into bicyclic analogues with only 'A' and 'B' rings, while 'A' ring substituents have been reviewed by Lown *et al.*[55]

It is thought that 'A' ring substitution (particularly oxy substituents at C8) might be involved in non-covalent binding with DNA through hydrogen bonding interactions with the minor groove.[56] An extensive study into 'A' ring modifications was carried out by Thurston *et al*, with the synthesis of five novel ring systems (**Figure 12**).[57]

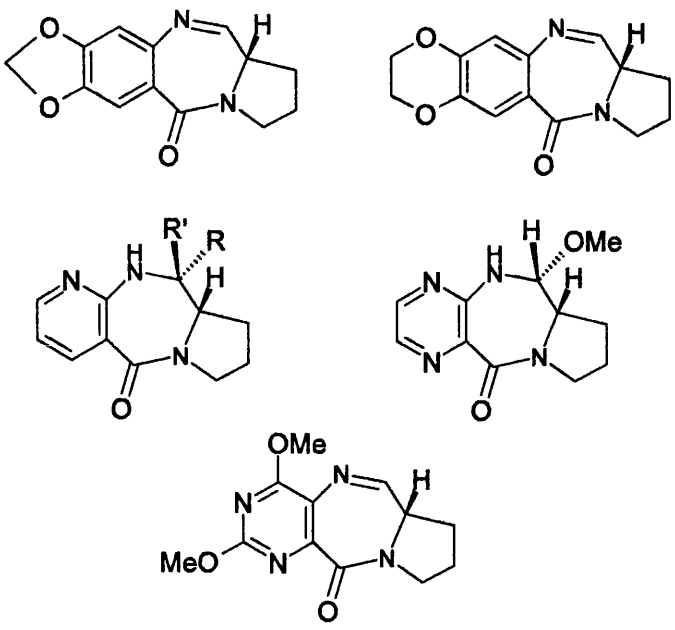


Figure 12: Five 'A' ring modifications[57]

This study clearly demonstrated the importance of the 'C' ring for cytotoxicity, as none of the 'A' ring modified compounds containing a saturated and unsubstituted 'C' ring had cytotoxicities even approaching the level of the three naturally occurring 'C' ring unsaturated/substituted compounds **(Figure 13)**.

Molecular modeling investigations into the modification of the 'C' ring suggested that C2 *endo* or C2 *exo* unsaturation leads to a flattening of the 'C' ring, and hence a more comfortable fit in the minor groove.[57] From the 'A' ring modifications it was found that an 'A' ring electron donating substituent is not a prerequisite for binding, but can influence binding affinity and cytotoxicity. Modeling of dioxazole and dioxazine analogues suggests that the extra rigidity introduced by the new rings may cause steric interactions with functional groups on the opposite minor groove walls, while C7 methoxy/C8 hydroxy substitution in DC-81 renders the molecule more flexible and allows for more pivoting – giving a better minor groove fit. Neither aza substitution of the 'A' ring or conversion of the ring to a five membered pyrazine improves either cytotoxicity or binding.

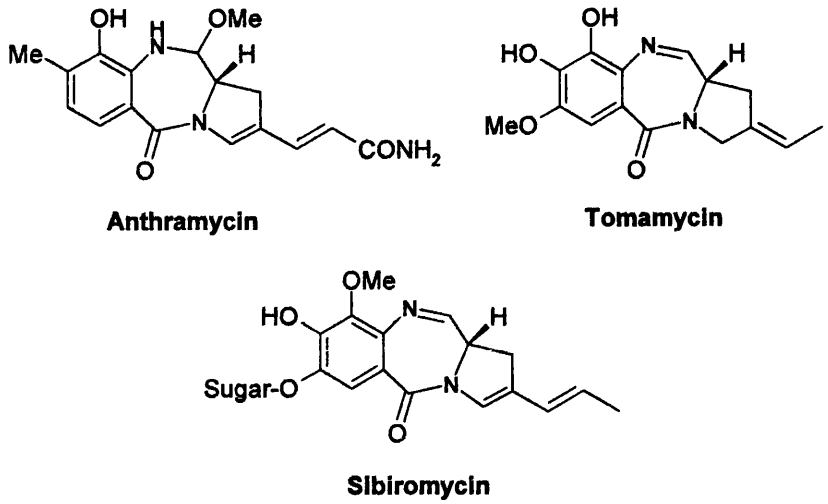


Figure 13: Three naturally occurring 'C' ring unsaturated/substituted compounds

Further investigations into A ring modification were conducted in 2001 when a new 7,8-methylenedioxy analogue of (+)-porothramycin B and its water soluble sodium bisulfite derivative (**Figure 14**) were shown to exhibit high cytotoxic activities against several tumour cell lines.[58] Investigation was carried out by Baraldi *et al* into synthesis of a pyrazolo[4,3-*e*]pyrrolo[1,2-a]diazepinone tricyclic system (**Figure 14**).[59]

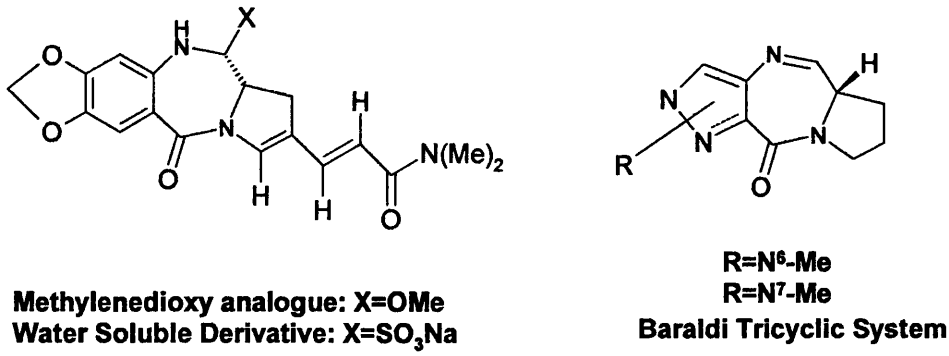


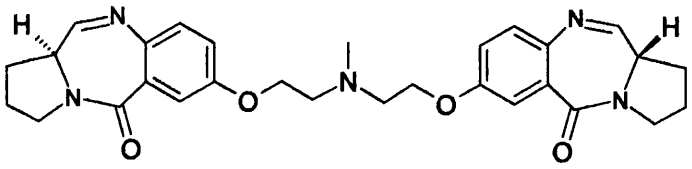
Figure 14: Bisulfite derivatives and the pyrazolo[4,3-e]pyrrolo[1,2-a]diazepinone tricyclic system[57, 59]

The rationale behind this design was to achieve a possibly higher binding affinity and modified sequence selectivity from the presence of potential new hydrogen bonds between the 'A' ring atoms and the DNA bases, as well as reduced cardiotoxicity due to the impossibility of the C9-quinone formation suggested to occur in the case of anthramycin. The findings from this study suggested that the presence of an unsubstituted imine function at N10-C11 is required for cytotoxicity, and for maximum cytotoxicity the pyrazole ring should preferably be substituted with a benzyl group at N7 and a methyl at C8. However, all of the analogues synthesized were significantly less potent than the PBD reference compounds. There was a relative loss of cytotoxicity overall when compared with DC-81 and this suggests that this type of modification to the 'A' ring may not be suitable for the enhancement of cytotoxicity. A significant difference in cytotoxicity between DC-81 (IC_{50} = 0.38mM) and tomamycin methyl ether (IC_{50} = 0.012mM) has also been noted elsewhere[59] and this, coupled with earlier 'C' ring observations, suggests that the 'C' ring of the PBDs is perhaps more important with regards to cytotoxicity.

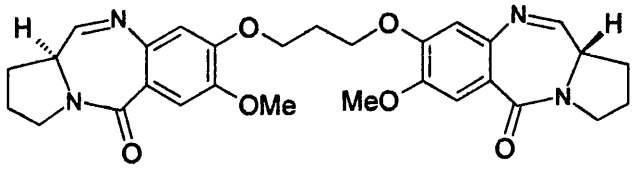
1.5.8: PBD Dimers

There has been increasing interest in bifunctional alkyating agents, forming inter-strand cross-links in the minor groove of DNA. These cross-linking agents are highly cytotoxic, possibly due to cross-links challenging cellular repair mechanisms[60] or agents targeting genes associated with cell growth.[61] Molecular modeling has shown covalent interaction of dimers within the minor groove to be twice that of the monomers, highlighting their difunctional character and the tolerance to the central linker.[62]

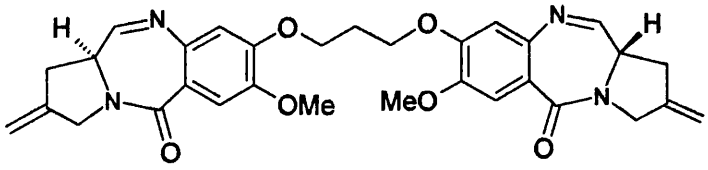
The first cross-linking agent was reported in 1991 by Suggs *et al.*[63] It consisted of two unsubstituted PBD units linked through their C7 positions and is shown in **Figure 15.**



1st PBD Dimer - Suggs *et al*, 1991



DSB-120 - Thurston *et al*, 1992



SJG-136 - Thurston *et al* - 2001

Figure 15: PBD Dimers[63,64,62]

The C7 linked dimer displayed poor DNA cross-linking activity and there was no evidence for cytotoxicity. In 1992 a C8 linked dimer of DC-81 known as DSB-120 was synthesized by Thurston *et al* (**Figure 15**).[64,65,66] DSB-120 showed efficient cross-linking activity and is a very potent cytotoxic agent.[67] Gel electrophoresis and thermal denaturation studies show that DSB-120 has a high affinity for DNA, forming irreversible symmetrical inter-strand cross-links between the spatially separated guanine N2 atoms on opposite DNA strands.[67] Due to the design of the duplex sequences however, it was not possible for the PBD to form an intra-strand link and so this possibility has not been previously investigated.

Molecular modeling and NMR studies indicate that DSB-120 covers a six base-pair region in the minor groove, showing a preference for base sequences 5'-Pu-GATC-py or 5'-Py-GATC-Pu.[67] A two step mechanism for the cross-linking is suggested and is shown schematically in **Figure 16.**[67]

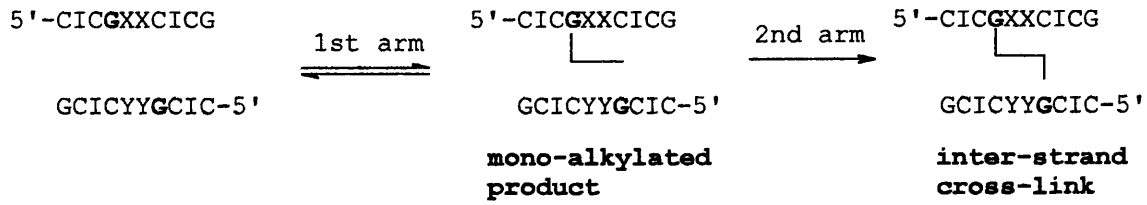


Figure 16: Suggested two-step cross-linking mechanism.
Reproduced from Thurston *et al*[67]

However, although DSB-120 was a very potent cytotoxic agent *in vitro*, it proved to have poor *in vivo* antitumour activity. Its low therapeutic index was thought to be partly due to reaction with molecules containing thiol, causing much reduced reaction at the tumour site.[69,72] As a result of these problems a new *exo*-unsaturated analogue called SJG-136 was synthesised (**Figure 15**).[62] The design of SJG-136 was based on the fact that PBD monomers with unsaturation at the C2 position are known to be more biologically potent - possibly a result of the C2-*exo* unsaturation leading to lower electrophilicity at the N10-C11 position.[72] SJG-136 has proved to be a highly efficient stabilising agent for double stranded calf thymus DNA, proving more than 10-fold more efficient than DSB-120. Molecular modeling shows that the SJG-136-DNA adduct maintains the B-form structure of the duplex, with little or no disruption of the secondary structure. Both SJG-136 and DSB-120 are well accommodated in the minor groove, with little exposure beyond the DNA periphery. This leads to the observed resistance to repair enzymes, which are thought to be reliant on tracing distortion or helical perturbation in DNA.[72] The calculated binding energies showed that PBDs with unsaturated 'C' rings formed more favourable adducts. Introduction of an *exo*-cyclic methylene at C2 causes stiffening of the 'C' ring, making it more planar and providing a better fit in the minor groove (**Figure 17**).[73]

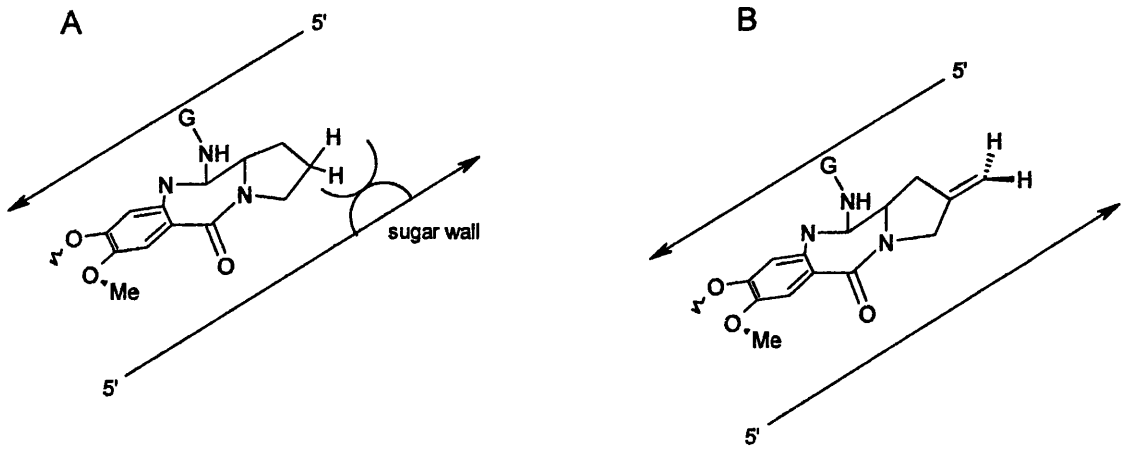
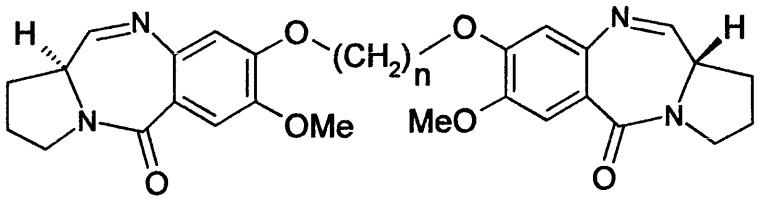


Figure 17: The fit of a PBD subunit within the minor groove
Reproduced from Thurston *et al*[73]

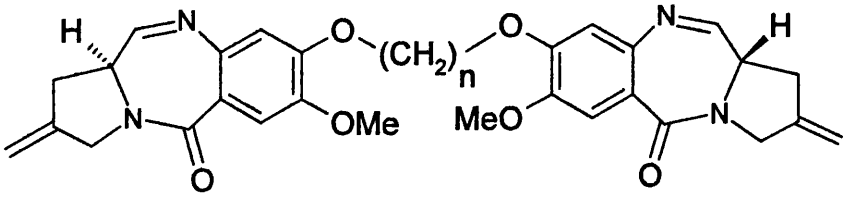
The linker itself was also found to make favourable contacts with the floor and walls of the minor groove. SJG-136 displays high *in vitro* cytotoxicity in cis-platin resistant cell lines, with an $IC_{50} = 2.25x10^{-5}$ M. This is the lowest IC_{50} value for any PBD monomer or dimer. The failure of DSB-120, as mentioned, was believed to be due to an interaction with glutathione or thiol containing proteins.[69] SJG-136 appears unaffected by high cellular GSH levels and so is expected to perform better. SJG-136 also exhibits a broad spectrum of *in vivo* antitumour activity, showing high efficacy in tumour xenograft models including melanoma, ovarian carcinoma, breast carcinoma or melanoma, colon carcinoma, promyelocytic leukaemia, and non-small cell lung carcinoma.[70] In addition, this versatile dimer, along with several analogues has been shown to possess antibacterial activity against MRSA, VRE and mycobacteria, with activity thought to stem once again from the covalent cross-linking of the DNA duplex.[71]

1.5.9: Effects of linker length in PBD dimers

Studies have been undertaken into the effect of linker length in dimers of both the saturated and the unsaturated type.[65,73] This involved syntheses of homologous series of C8 linked bifunctional DNA alkylating agents with varying linker length, and study of their cross-linking efficiency and cytotoxicity (**Figure 18**).[65,73]



3a - 3d n=3-6



4a n=3, **4b** n=5

Figure 18: Series of C8-linked bifunctional DNA alkylating agents with varying linker lengths[65, 73,74]

Linkers with an odd number of carbon atoms (n=3, n=5) were found to have a similar cross-linking efficiency, while those with an even carbon chain were up to 18-fold less efficient. This reflects the difference in ability of the compounds to stabilise the DNA double helix coil transitions. In the saturated series **3a-3c** the three carbon linker **3a** (DSB-120) appeared to bind more tightly to double stranded DNA than the five carbon (**3c**) compound. This was in contrast to the unsaturated **4a-4c** series, where lengthening of the $-O-(CH_2)_n-O-$ linker from three to five effected an unexpected enhancement of DNA reactivity and *in vitro* cytotoxicity. The cross-link potency is increased around 10-fold for **4c** relative to **4a**, in stark contrast to the **3a-3c** series, where the increase of the carbon chain from three to five somewhat reduced the cross-linking potency. *In vitro* cytotoxicity was also significantly higher in **4b** than for the three-carbon linker **4a**. Enhancement of activity in individual cell lines ranged from around 30-fold (melanoma) to greater than 3000-fold in ovarian cancer cells. This increase in activity was again

unmarked in the **3a-3c** series. Molecular modeling has helped to explain this behaviour in terms of a superior isohelical fit of the C2-*exo*-methylene substituted 'C' rings within the minor groove, such that the steric crowding between the groove walls and the ring C2 proton of the saturated analogues is avoided.[73]

In a related study a series of radio-labelled oligonucleotides were synthesised to enable the comparison of reactivity and cross-linking efficiency within the varying linker lengths.[72] The oligonucleotides synthesised are shown in **Table 1.**

Oligonucleotide Structure		General Structure
Oligo-1	5'-(CIC-GATC-ICG)$_2$	Py-G-Pu
Oligo-2	5'-(CIC-GTAC-ICG)$_2$	Py-G-Py
Oligo-3	5'-(TATA-GATC-TATA)$_2$	Pu-G-Pu
Oligo-4	5'-TATA-GATTC-TATA 3'-ATAT-CTAAG-ATAT	Pu-G-Pu

Table 1: Series of radio-labelled oligonucleotides synthesised for comparison of reactivity and cross-linking efficiency for varying linker lengths[74]

Denaturing polyacrylmide gel electrophoresis techniques[74] showed Oligo 1 to be cross linked around 2 fold more efficiently than Oligo 2, and Oligo 3 showed a preference for cross-linking greater than Oligo 1 confirming the 5'Pu-G-Pu > 5'-Py-G-Py ranking predicted by DNA footprinting and molecular modeling. This behaviour is due to favourable hydrogen bonding between the N10-H10 proton of each bound PBD moiety and the N3 ring nitrogen of the 3' side adjacent adenines (**Figure 19**).[67]

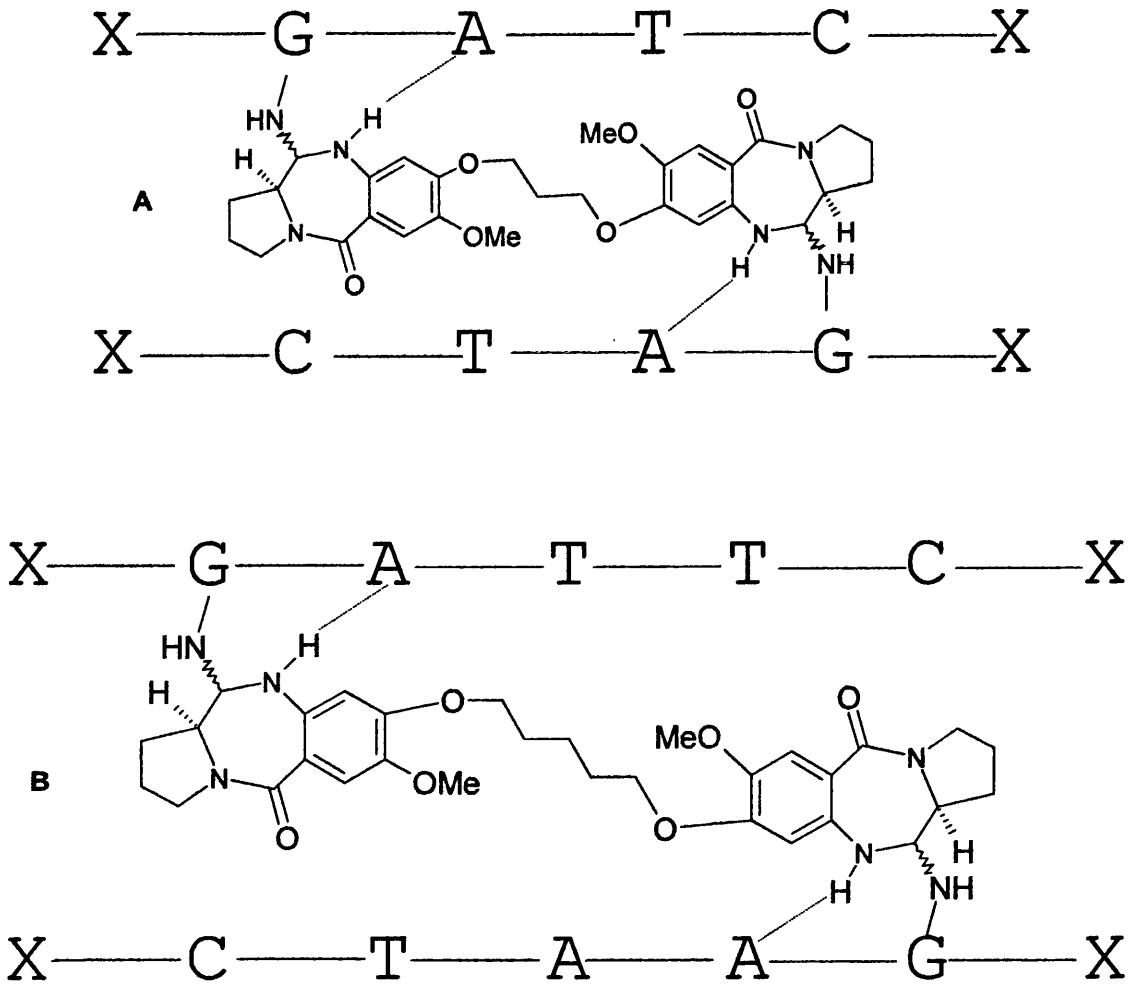


Figure 19: Sequence selectivity due to covalent bonds and hydrogen bonds
Reproduced from Thurston *et al*[73]

As in the earlier study, compounds with an odd numbered linker (n=3, n=5) proved to have a much greater cross-linking efficiency than the even. There was significant correlation between DNA binding affinity, cross-linking efficiency and cytotoxicity, and the three carbon linked PBD proved to be 300-fold more efficient at cross-linking DNA than the clinically used minor groove binding cross-linking agent melphalan.

It was found that the n = 3 homologue preferred a 5'-GATC- sequence but could not tolerate an additional base pair between the two guanines, whereas the n = 5 dimer

could cross-link both the 5'-GATC- site and the longer 5'-GAATC- site by adoption of a folded linker conformation. Molecular modeling showed that only the n=3 and n=5 dimers were able to form energetically favourable adducts, with the reactivity dictated by the spatial distribution of the PBD subunits. In the case of the even homologues the unfavourable energy terms are a consequence of poor geometry in the diether linkage, leading to a distortion of the host DNA and a poor conformation adopted by the bound ligands.[72] Models of the n=5 dimer showed the linker group held snugly against the hydrophobic walls of the DNA minor groove in the case of the longer sequence, but partly displaced away from the helix and compacted by internal conformational rotation to achieve cross-linking with the shorter 5'-GATC- tract.[73]

As with the mono-alkylating PBDs, it is believed that the stereochemistry at the PBD-C11 link preferably exists in the 'S' configuration.[38] Attempts at molecular model refinement with 'R' stereochemistry lead to destabilization of the cross-linked adduct resulting in disruption of the base pair associated with the covalently modified guanine. This caused the complimentary base cytosine to uncouple from the stacked helix and become displaced into the minor groove with an energetically unfavourable distortion of the local DNA backbone.[67] Modeling with the 'S,S' stereochemistry produced stable adducts which retained general B-type DNA integrity, and accommodated all NOE data. This is reminiscent of the PBD-C11 'S' stereochemistry of a 2:1 anthramycin-DNA complex confirmed by X-ray crystallography.[41] For this complex, formation of 'R' stereochemistry again lead to distortion.[41]

1.5.10: Summary

The PBDs are an interesting and versatile group of DNA alkylating agents with relatively simple structures. They react covalently with the NH_2 group of a guanine DNA base and thereby inhibit DNA transcription. PBD dimers have been well researched and are capable of forming inter-strand cross-links with duplex DNA. To date no previous evidence has been found for either mono-adducts or for the intra-strand cross-linking mechanism discussed in this project.

1.6: The Cyclopropapyrroloindoles (CPIs)

Cyclopropapyrroloindoles are a class of drugs containing the DNA alkylating cyclopropa[*c*]pyrrolo[3,2-*e*]-indole-4[5*H*]one subunit. The entire class shows exceptional antitumour activity, deriving their biological effects through sequence selective DNA alkylation.[75-77]

1.6.1: History

(+)-CC-1065 was the first CPI to be discovered, isolated from *Streptomyces zelenis* by the Upjohn Company in 1978.[78] It consists of three subunits with the CPI moiety in subunit 'A' (**Figure 20**).

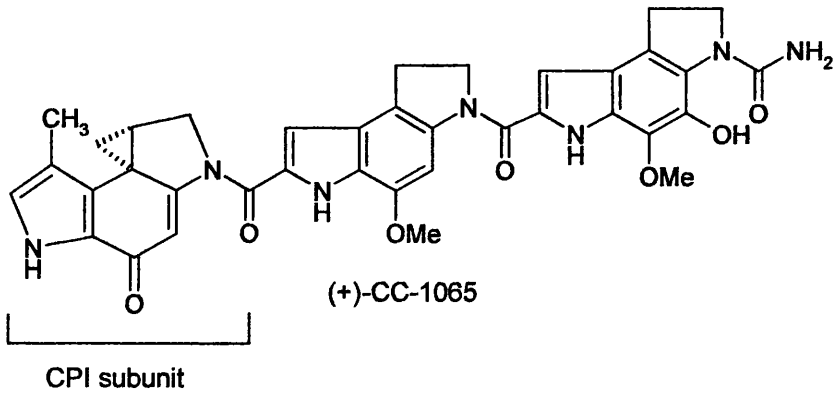


Figure 20: (+)-CC-1065, highlighting the CPI subunit

The curved shape of this drug facilitates association with the minor groove of B-form duplex DNA and due to its DNA binding ability it was found to be a potent antitumour antibiotic.[79,80] It was discovered that the CPI subunit alone contained enough information to encode for sequence selectivity and produce bending of DNA[81] while

additional substituents on the non-alkylating subunits serve to fine-tune sequence selectivity and produce DNA winding.[81,82]

Despite the high potency of (+)-CC-1065, development of this drug was abruptly terminated on the discovery of an unusual hepatoxicity that caused delayed lethality in mice at sub-therapeutic doses.[83] It was later confirmed that the cause of the delayed death was associated with the ethano bridges of the non-alkylating subunits, and these were also proposed to be responsible for the winding of the DNA in the drug-DNA adducts.[84] Accordingly, new analogues of the drug including adozelesin, bizelesin and carzelesin (**Figure 21**) were synthesised in order to attempt to combat the delayed lethality problems associated with (+)-CC-1065.[85,86] These compounds will be discussed in further detail in a subsequent section.

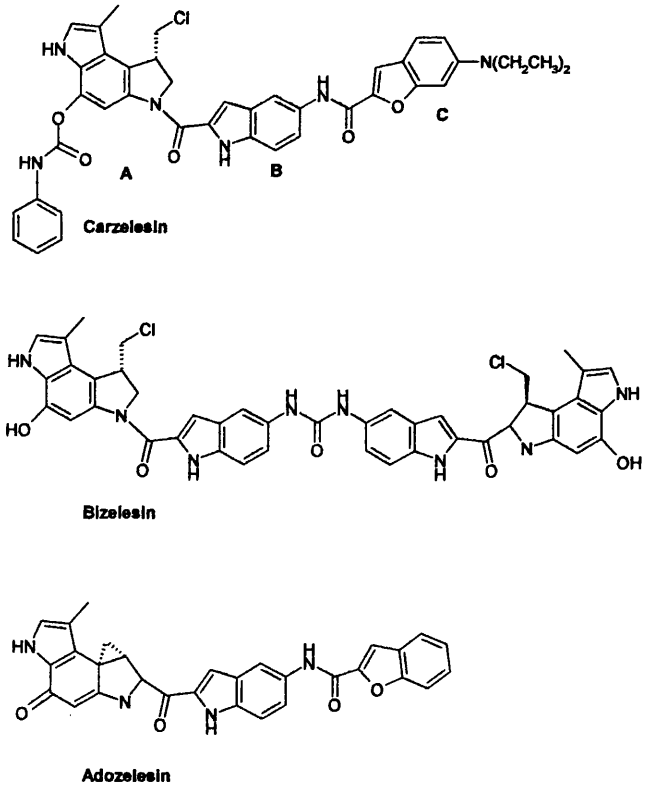


Figure 21: Chemical structures of carzelesin, bizelesin and adozelesin

1.6.2: Mode of Action

In analogy with the DNA minor groove binders netropsin and distamycin, (+)-CC-1065 exhibits a right handed twist along its axis.[87] The hydrophobic groups are situated on the concave side of the molecule and experience favourable interactions with the minor groove while the hydrophilic groups are found on the convex side of the molecule.

Early research into the sequence specificity and mode of action of (+)-CC-1065 concluded that cytotoxic effects are mediated *via* DNA synthesis[79] and that the drug binds covalently to DNA within the minor groove.[88]

(+)-CC-1065 reacts with double stranded DNA through the N3 position of adenine to form a covalent adduct that overlaps a five base pair region in the minor groove.[88-90] It shows a high sequence selectivity for AT regions, reacting preferably with 5'-PuNTTA* and 5'-AAAAA*,[91] and showing an absolute requirement for an adenine at the 3' end.[91] The reaction of the drug with DNA is thought to be a multi-step process with non-covalent interaction through hydrophobic and *Van der Waals* forces between the 'B' and 'C' units with the minor groove preceding covalent bonding with the alkylating subunit 'A'.[92,81]

(+)-CC-1065 and adozelesin both exist as cyclopropyl drugs, and the mechanism of reaction proceeds *via* nucleophilic attack on the cyclopropyl ring by the N3 of an adenine nucleotide to form a DNA adduct with 'S' stereochemistry at the C4a centre, as shown in **Figure 22.** Bizelesin exists in a slightly less active chloro-methane form, which requires activation prior to reaction with DNA. Activation proceeds *via* deprotonation of the phenol group in aqueous conditions as shown in **Figure 22** and is hastened by addition of sodium carbonate. Carzelesin exists as an inactive prodrug as shown in **Figure 21.** Activation is effected by hydrolysis of the phenylurethane substituent to a chloromethyl phenol, followed by a ring closing step to produce the cyclopropyl moiety required to alkylate DNA as discussed in relation to (+)-CC-1065 and adozelesin.

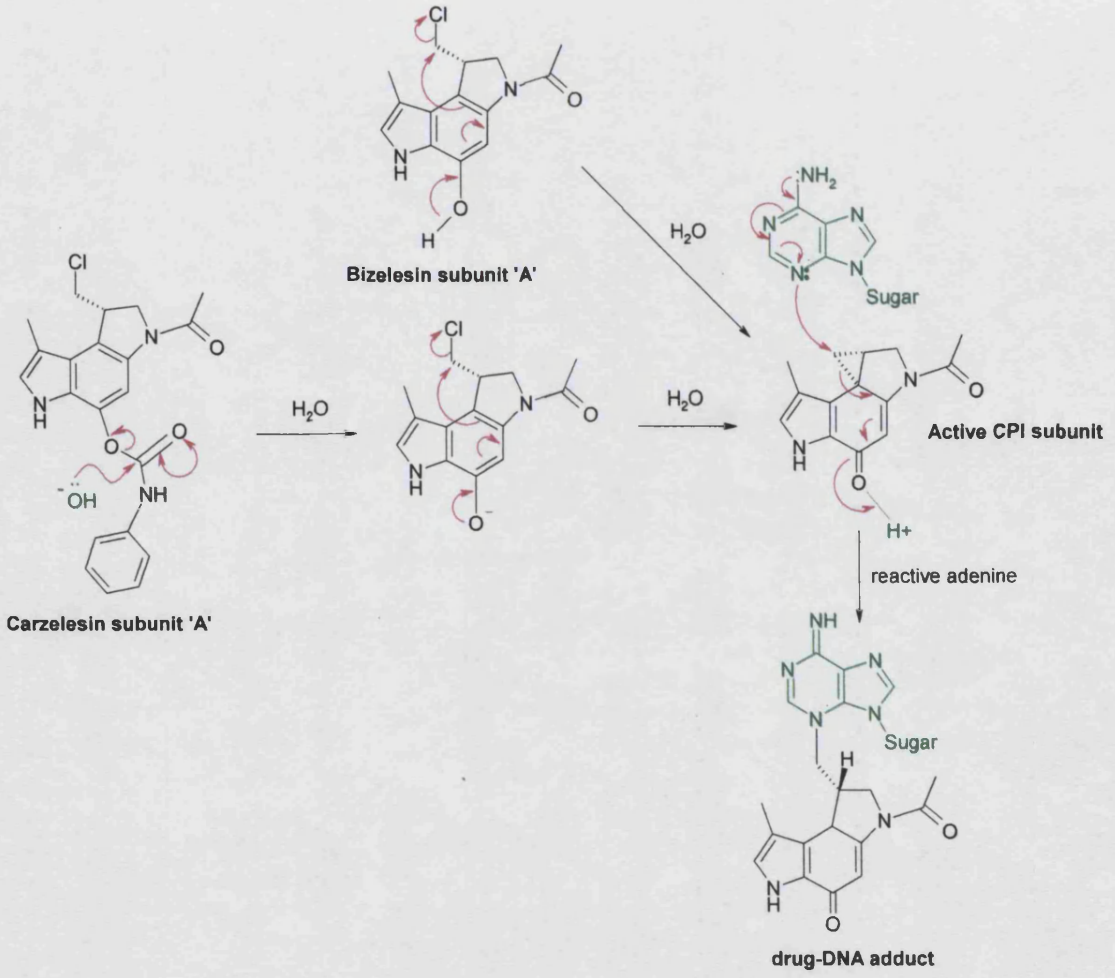


Figure 22: Activation of carzelesin and bizelesin followed by general CPI reaction mechanism

Important structural attributes of (+)-CC-1065 were investigated by Warpehoski *et al.*[93] A series of (+)-CC-1065 analogues were synthesised in order to investigate structure-activity relationships and it was found that;

1) An acyl substituent located on the pyrrolidine nitrogen of the CPI optimises reactivity to acid catalysed nucleophilic attack,

2) The ring size of the 'B' subunit is important for DNA binding – a five membered ring heteroaroyl substituent attached to the CPI permits binding to the DNA helix

3) Extension of the amide-linked indole chain enhances DNA interaction.

The delayed hepatotoxicity of (+)-CC-1065 discussed previously is thought to be caused by the ethylene bridge moieties of the 'B' and 'C' subunits, which inhibit helicase-mediated unwinding, ultimately causing over-winding of the DNA helix.[83,94,75]

In order to investigate non-covalent versus covalent interactions, analogues of (+)-CC-1065 were synthesised by Hurley *et al*[81] (**Figure 23**). These were compared to (+)-CC-1065 with respect to sequence specificity and biological potency of the covalent and non-covalent interactions.

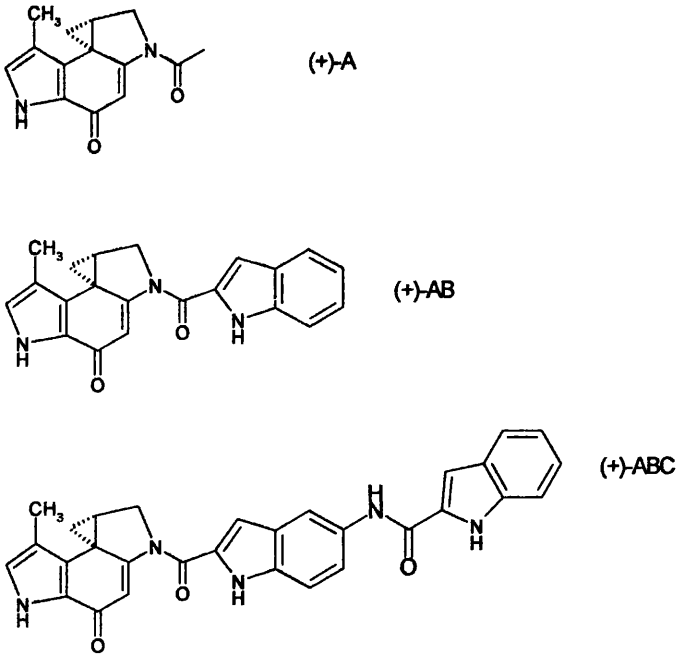


Figure 23: Analogues of (+)-CC-1065 synthesised by Hurley *et al*[81]

The results of this investigation confirmed that biological potency was linked to the alkylation of DNA. It was found that an analogue of (+)-ABC minus the cyclopropyl ring did not bind to DNA, therefore implicating the cyclopropyl ring in the binding mechanism. Studies using (+)-A and (+)-AB showed that the 'B' and 'C' subunits increased the rate of alkylation at specific adenines. All analogues showed an equal sequence specificity in adduct formation, suggesting that it was the DNA base sequence that was responsible for the reactivity of adenines to the CPI subunit, as opposed to interactions with the 'B' and 'C' subunits. These subunits were instead suggested to have a role in the stabilization of non-covalent bonding and possibly the fine tuning of the sequence selectivity prior to alkylation.

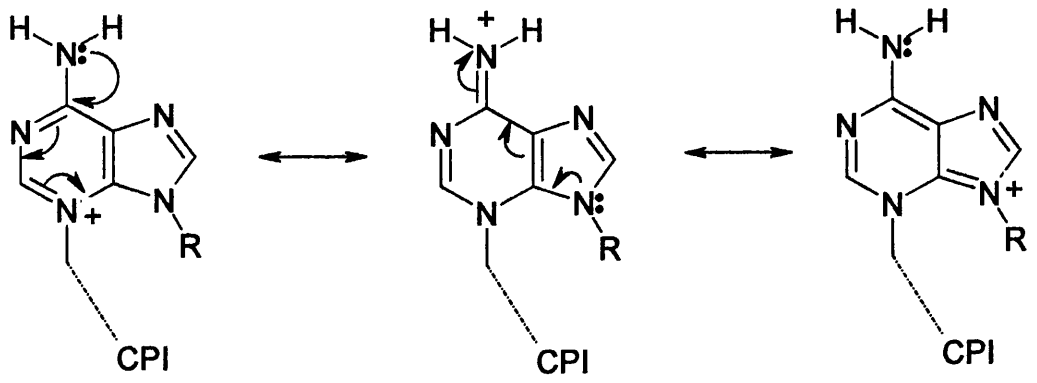
This view was challenged in 1990 when Boger and co-workers proved that biological activity was not dependent on electrophilic reactivity.[95,96,97] It was found in these studies that the rate of DNA alkylation and the cytotoxic potency were inverse to the electrophilic reactivity. The 'B' and 'C' subunits were now thought to play a large role

in sequence selectivity through non-covalent binding, preferentially in an AT-rich minor groove environment.

Analysis of the 5'*d*(CGCGGAGTTA*GG)-(+)-CC-1065 adduct by NMR[98] using ^{17}O-labelled water and phosphate resulted in the incorporation of two water molecules into the alkylation reaction pathway. The first bridges the *O*-phenolic proton of the CPI subunit to the anionic oxygen of the phosphate two bases in the 3' direction of the thymine residue of the covalently modified base pair. The second is found to be hydrogen bonded to the alkylated adenine. Both molecules are thought to be involved in the mechanism of covalent bond formation due to their prolonged dwelling times. The involvement is thought to involve acid catalysed activation of the CPI subunit in the reaction of N3 of adenine.

Studies have been carried out into the extent of drug induced DNA strand damage by (+)-CC-1065 using adducts with the Simian virus 40 (SV40) DNA molecule.[99] It was thought that the accessibility of cellular DNA to the action of (+)-CC-1065 might be limited by the protein components of cellular chromatin, or by other factors such as cell and nuclear compartmentalization. It was found that the pattern of (+)-CC-1065 adduct formation in SV40 infected BSC-1 cells is almost the same as that observed with the purified SV40 DNA molecule. However there is a requirement for higher drug levels in the binding to infected cells, and it is thought that uptake parameters may be the reason for this reduced efficiency in adduct formation.

In CPI-DNA adducts the DNA duplexes are extremely stable and it was originally proposed that due to this stability the adenine existed in a neutral state with the possibility that an imine form was present. However, further calculations[100] showed that a positively charged form could also exist in DNA, and ^{1}H and ^{15}N NMR of the (+)-CC-1065-(^{15}N6-adenine)-DNA adduct settled upon a final structure in which the adduct existed in a doubly protonated form, with the charge located over the entire modified adenine (**Figure 24**).[101,102]



Major species

Figure 24: Mesomers of adenine in its amino form, showing the major species

1.6.3: Inter-strand cross-links of (+)-CC-1065

Despite the fact that (+)-CC-1065 is established to be a mono-alkylator hypotheses have been formulated around the possibility of enzyme-activated inter-strand cross-link formation with this drug. It was thought that the drug might undergo metabolic activation by cellular enzymes and so be able to form an additional covalent bond with DNA. Experimentation revealed that inter-strand cross-linking was indeed observed for (+)-CC-1065 in the cellular system.[103] No cross-linking was seen when the drug was added directly to cellular lysates with inactivated enzymes, supporting the theory of enzymatic activation. The reactive part of the parent drug is the cyclopropyl ring. This is the site of monoalkylation with the second site required for a cross-link generated by activation of one of the two available *o*-methoxy phenol groups. *o*-Methoxy-phenol groups are present in a variety of biologically active compounds and a possible pathway of metabolic activation is shown in **Figure 25.** A second mechanism was also proposed in which the activated indole moiety of (+)-CC-1065 reacts with 5,6-dihydroxy or 5,6-quinone intermediates. This reaction pathway has been shown for melanin precursors.[103]

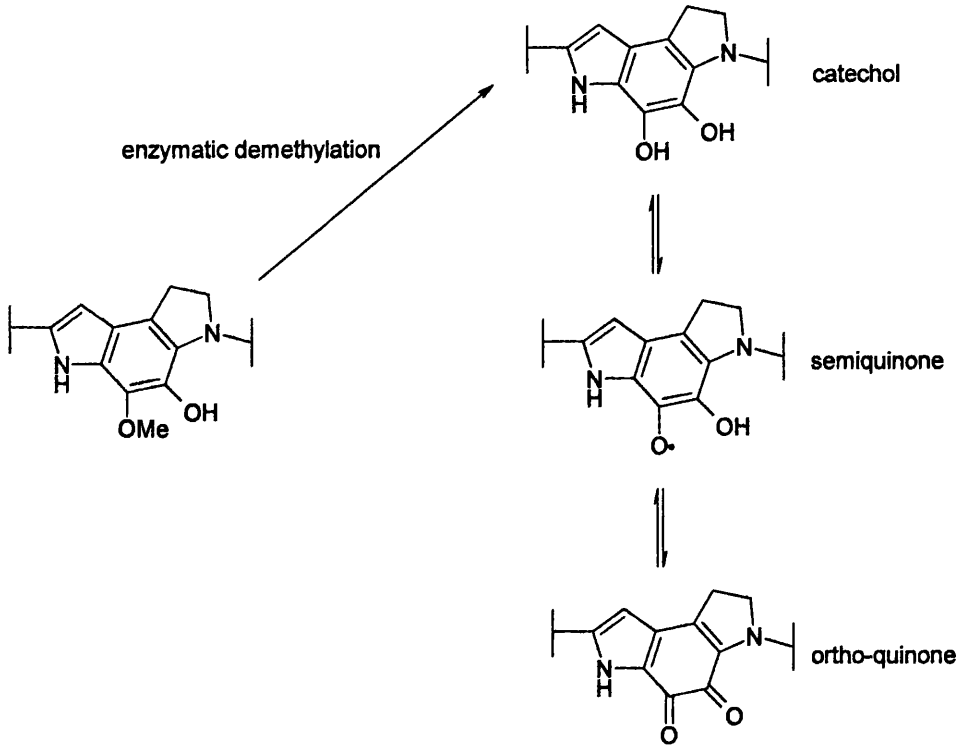


Figure 25: Possible pathway of metabolic activation for *o*-methoxy-phenol groups[103]

1.6.4: Adozelesin

Adozelesin was the first of the new generation analogues and is shown in **Figure 21**; like (+)-CC-1065 it is overall right-handed and mimics the pitch of B-form DNA. This characteristic allows it to fit snugly in the minor groove of DNA with a high recognition for its binding site, showing a sequence preference of 5'-Pu,Py/Pu,TTA* (where the asterisk shows the site of covalent modification).[91] The adozelesin duplex adduct was reported by Cameron and Thompson;[104] experiments were performed using a single palindromic DNA duplex containing two identical adozelesin binding sites, namely 5'*d*($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$)$_2$. The adozelesin residues bound to the A12 nucleotide in an edge-on orientation in the minor groove, and traditional Watson-Crick base pairing was maintained. Significant hydrogen bonding was found to exist between phenolic protons and the phosphate backbone, shown by the chemical shift of the phenol of the CPI subunit.[104] A strong hydrogen bond was also formed between the

NH proton of the amide linker between the indole and benzofuran subunits and a thymine (T10) on the modified strand (**Figure 26**). This hydrogen bonding is much stronger than the hydrogen bonding between the phenolic protons and the phosphate backbone but surprisingly causes much less detectable distortion throughout the molecule. However some distortion was seen around the central guanine (G6), which appeared to have adopted a less favoured C3'-*endo* conformation whilst the other 13 base pairs remain in the more commonly observed *anti* configuration. This provides additional hydrogen bonding opportunities between the ligand and the duplex.

As with (+)-CC-1065, adozelesin owes its biological activity to the ability to block DNA replication. It has been suggested that this inhibition occurs *via* a trans-acting mechanism through cellular damage response pathways or checkpoints, rather than by directly blocking DNA replication, and subsequent experimentation[105] has shown that the replication is inhibited through the inactivation of human replication protein (RPA) – a major eukaryotic single-strand DNA binding protein required for DNA replication, repair and recombination.[106]

Adozelesin had high potency and showed good preclinical activity against murine and human tumour models.[107,108] As discussed above studies on the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct[104] showed significant hydrogen bonding that suggested that the binding mechanism for adozelesin was significantly different to (+)-CC-1065, possibly resulting in an inability of adozelesin to overwind DNA, and this drug has shown no sign of the delayed toxicity that halted the development of (+)-CC-1065. However, in clinical trials the drug proved to have only marginal efficacy, showing dose-limiting toxicities of myelosuppression with thrombocytopenia and leukopenia.[109]

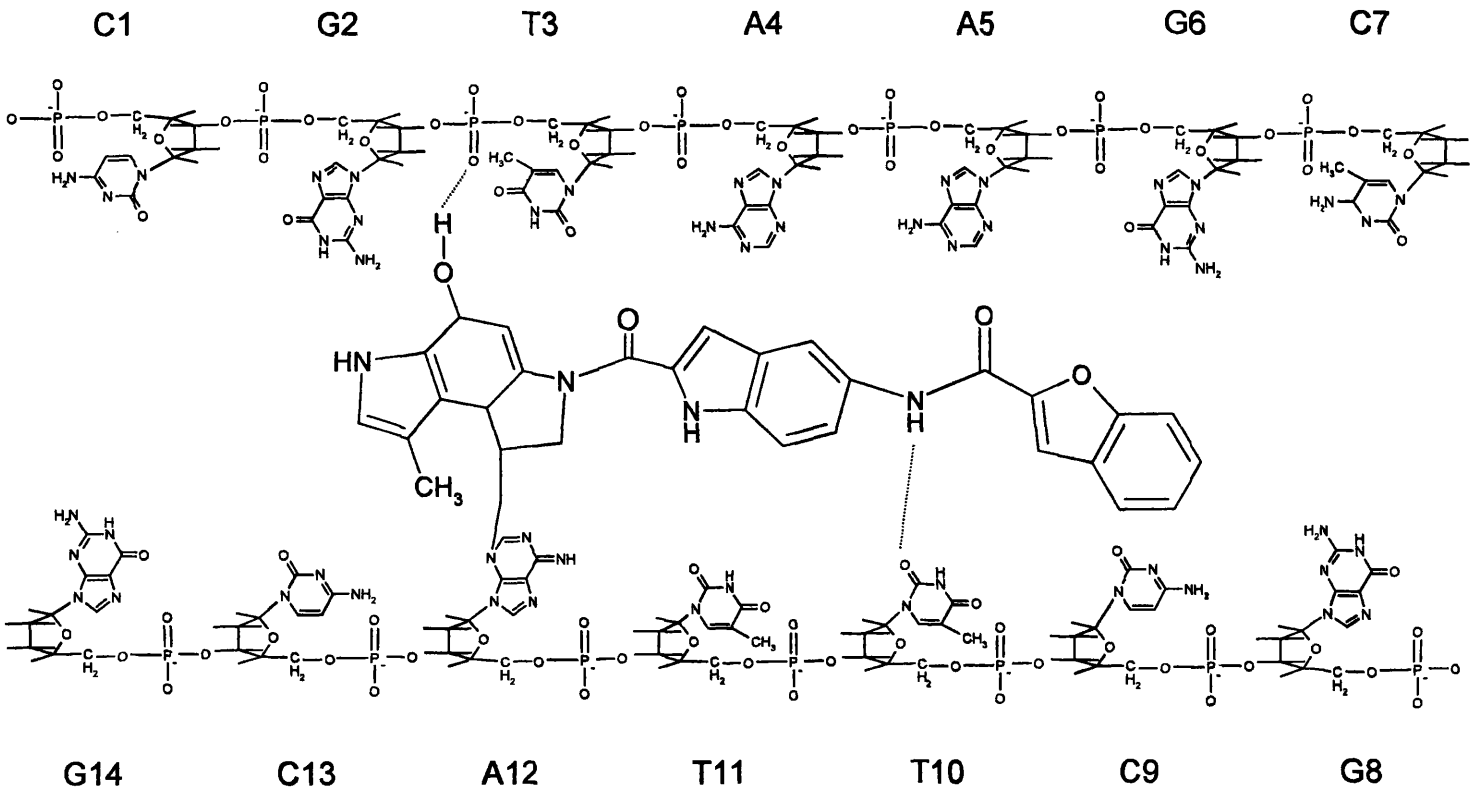
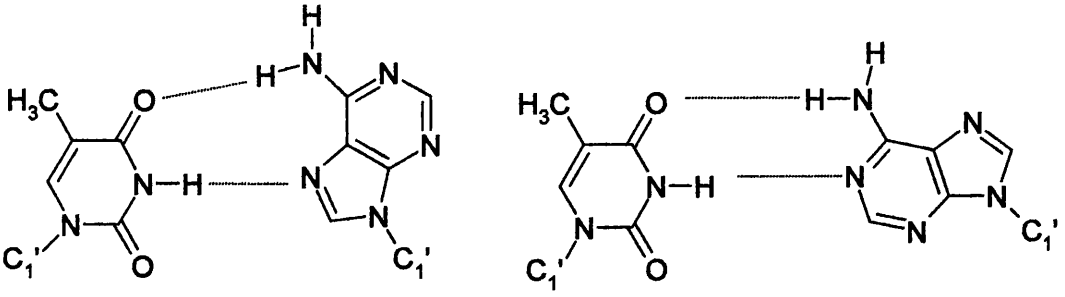


Figure 26: Schematic representation of adozelesin accommodation in the minor groove of B-form DNA, showing potential hydrogen bonds between the drug and the DNA bases.

1.6.5: Bizelesin

Bizelesin is a synthetically derived bifunctional DNA cross-linking agent designed and synthesised by Upjohn scientists.[110] It is an analogue of (+)-CC-1065, containing two open cyclopropyl ring analogues of the CPI subunit, which are connected by a rigid linker consisting of two indole units joined by a ureadiyl linker (see **Figure 21**). The drug once again reacts through the N3 position of adenine in the minor groove, to form inter-strand cross-links with adenine residues, generally six base pairs apart.[111] The sequence specificity of bizelesin was studied using restriction enzyme fragments.[112] In this experiment a higher than expected proportion of cross-linked *vs* mono-alkylated adducts was seen, and this was rationalized on the basis that the thermodynamic stability of the cross-linked adduct is higher relative to the mono-alkylated species. Bizelesin GC tolerance was also twice as high at the cross-linking sites than for the mono-alkylated species. This is due to increased reactivity at normally non-reactive sites, caused by covalent immobilisation of the drug leading to close proximity of the second alkylation arm. Where bizelesin mono-alkylation occurred the drug appears to show an increased sequence selectivity relative to the mono-alkylating analogues.[111] The most preferred sequence for bizelesin is known to be a six base-pair region – 5'TAATTA*-3',[113,114] mimicking the preference of the mono-alkylating drug (+)-CC-1065. On reaction of bizelesin with this sequence two adduct conformations are present, both with unusual base pairing within the central AT bases.[115] The adduct exists as 40% with base pairs in an open conformation, while the remaining 60% is Hoogsteen base-paired as shown in **Figure 27**. The origin of this unusual base pairing is possibly due to low base pair stability in this particular sequence. It is possible that bizelesin traps out and stabilises Hoogsteen and open base pairs prior to cross-linking.



Hoogsteen T-A base pairing

Watson-Crick T-A base pairing

Figure 27: Hoogsteen and Watson-Crick base pairing in DNA bases

The second preferred cross-link sequence is 5'-TTTTTA-3', and binding of (+)-CC-1065 and bizelesin to this sequence was examined by high field NMR and PAGE experiments using a 5'*d*(CGTTTTTACG).5'*d*(CGTAAAAACG) duplex.[116,117] DNA containing 'A' tracts has been shown by techniques such as gel electrophoresis, electron microscopy and X-ray crystallography to contain an unusually narrow minor groove which is intrinsically bent.[118-126] On reaction with (+)-CC-1065 the bent duplex is observed. However, on reaction with bizelesin the bending is eliminated. This supports the theory for origin of 'A' tract bending shown by UV and CD spectroscopy,[127,128] in which it is suggested that both straight and bent forms exist in equilibrium, and the degree of bending depends on the position of this equilibrium. In this experiment it has been observed that (+)-CC-1065 freezes out the bent conformer while bizelesin entraps the straight one. On formation of the inter-strand cross-link with bizelesin A8 and A18 are modified **(Figure 28).**[129] Relative Molecular Dynamics studies performed on the linker within the bizelesin-5'-TTTTTA*-3' adduct also show the linker oscillating between two conformations, unable to find hydrogen bond acceptors or donators in either position.

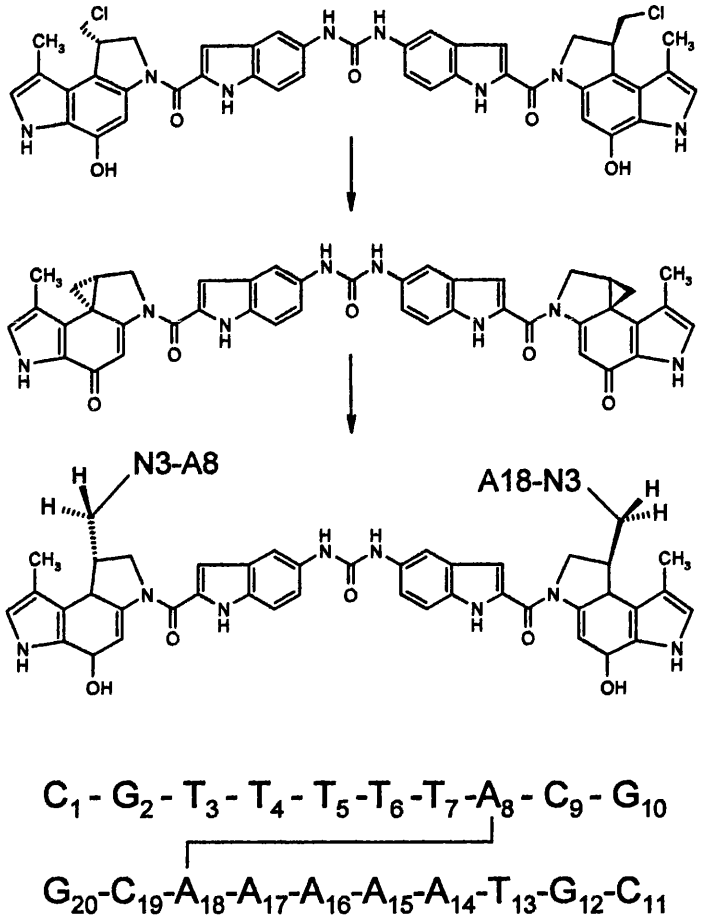


Figure 28: Formation of inter-strand cross-link *via* modification of A8 and A18[129]

As discussed in section 1.2.2, bizelesin is supplied as a less active chloro-methane species and undergoes activation prior to reaction with DNA. Post-activation modification of the DNA duplex by bizelesin occurs by attachment once again through position N3 of adenine and the ligand is again orientated end on in the minor grove. In this 'A' tract adduct there has been no evidence for unusual conformations - Watson-Crick base pairing is maintained throughout. This supports the above-mentioned hypothesis for low base pair stability, as the 'A' tract sequence is relatively stable compared to the 5'-TAATTA- duplex. As noted with (+)-CC-1065, in the case of the 'A' tract sequence both covalently modified adenine residues exist in a doubly protonated form. The trapping of straight DNA is potentially useful as a probe to evaluate the importance of intrinsic bending or protein induced bending in transcriptional or replication elements.

Molecular modeling and gel electrophoresis experiments have indicated that bizelesin is most suited to cross linking two adenines spaced six base pairs apart.[129] However, a third cross-link type has also been reported in which bizelesin was noted to span a total of seven base pairs, sometimes in preference when a six base-pair site was available.[129] Experiments were carried out using both 5'-TTAATTA-3' and 5'-TTAGTTA-3' sequences, and it is the presence of the –GC- base pairs in the centre of the sequence that provide the key to seven base pairing preferences. In the first sequence the six base pair site is favoured whereas in the sequence containing the GC base pair, distortion around the covalently modified adenine leads to a reduction in distance thus allowing a highly favoured seven base pair extended cross-link site. The covalently modified adenine is pulled toward the centre of the duplex resulting in a high degree of propeller twist and loss of hydrogen bonding into the phosphate backbone. The *exo*cyclic 2-amino group of guanine in the minor groove facilitates this distortion by the formation of a stable hydrogen bond between the carbonyl of the ureadyl linker and the *exo* amine group of guanine. Replacement of this central guanine group by inosine leads to the formation of approximately equal amounts of six and seven base pair links. In view of the association of the central guanine with sites on the linker, it would be possible in theory to change linkers in order to change sequence specificity and reactivity.

Although bizelesin alkylates DNA with a similar sequence preference to adozelesin, it has been shown that different biological responses are affected by this induced damage.[130] Both bizelesin and adozelesin inhibit DNA replication but while adozelesin induces this affect by a decrease in functional RPA[105] bizelesin inhibits replication through induction of a replication inhibitor.[131] At equi-toxic concentrations adozelesin and bizelesin induce different cellular responses, with adozelesin inducing an 'S phase' slowdown preceding apoptosis while bizelesin induces cell cycle arrest in the 'G2-M phase' through activation of p53 and p21 pathways and cellular senescence. These differences are thought to be due to the type of lesion and the extent of the damage – adozelesin may more extensively damage DNA and cause apoptosis, while the fewer but more cytotoxic bizelesin double strand breaks lead to cell cycle arrest.[130]

As with (+)-CC-1065 experiments were carried out into the effects of bizelesin on genomic DNA of BSC-1 cells as well as intra-cellular and purified Simian virus 40

(SV40) DNA.[132] It was shown that drug bonding sites were fewer for bizelesin than for (+)-CC-1065[99] but similar regions of the SV40 genome were affected by both drugs. Lesions induced by bizelesin also proved 100 times more cytotoxic than those induced by (+)-CC-1065. Similar regions of the genome were damaged in both intra-cellular and purified SV40 DNA systems, confirming that the chromatin structure of the intra-cellular DNA does not alter the localization of the bizelesin bonding. Once again however, the drug concentrations required to produce damage in the intra-cellular DNA are higher than for purified SV40 systems, suggesting as before, a limited accessibility into the cell.

Investigation has been carried out into the possibility of region-specific damage to 'AT islands' by bizelesin.[133] Long range *in silico* analysis shows that at the genomic level the bizelesin binding motif is found on average 2.8 times per 250 base pairs. However there is a regular occurrence of AT-rich hot spots appearing approximately every 10^6 base pairs, in which up to 99 binding sites are found per 250 base pairs. These are known as 'AT islands' and in models they have proved to be 100-fold more active than non-AT islands and relative to these AT rich sites the short AT tracts are unlikely to be significant as a bizelesin target. One attribute of an AT island domain is a high Matrix Associated Region (MAR) potential. MAR domains are a specialized loci that anchor cellular DNA to the nuclear matrix[134-137] and are the site of DNA replication and transcription.[134,136-140] Therefore bizelesin-induced damage to MAR-like AT islands may explain the observed specific inhibition of replication.[141,131] There are many potential therapeutic benefits associated with the targeting of AT islands. Certain abnormal AT islands are thought to contribute to tumourigenic phenotypes[142] and AT island targeting may offer increased selectivity. Also, tumour cells divide rapidly and so should be preferentially eliminated by highly localized lesions in replication-related regions such as MAR-like AT islands. Adozelesin also shows some preference for AT islands but there is substantial binding outside these regions. This shows that sequence specificity alone is not sufficient to induce region specificity. The oligopeptide antibiotic **tallimustine** exhibits a preference for 5'-TTTTGPu-3' sites, which are distributed nearly randomly throughout the genome so the drug displays a lack of region specificity. These cases when compared with bizelesin highlight the importance of both sequence specificity and genomic regions targeted in the design of anti-cancer drugs that limit the adverse effects of current treatments.

1.6.6: Summary

The CPIs are a family of minor groove binders that alkylate DNA by reaction at the CPI-C4 position to form a covalent bond with the N3 of reactive adenine. Various analogues exist including (+)-CC-1065, carzelesin, adozelesin and bizelesin. The drugs are overall right-handed and mimic the pitch of B-form DNA. They fit snugly into the minor groove and show exceptional antitumour activity, mediating their cytotoxic effects *via* DNA replication. A high sequence selectivity for AT regions is shown, with a preference for 5'-PuNTTA* and 5'-AAAAA* sequences observed. Formation of a 5'*d*(CGTAATTACG)$_2$-bizelesin adduct resulted in two conformations of the DNA duplex. The first exhibited open base pairs while in the second adduct the central AT step was found to be Hoogsteened. It was thought that bizelesin could trap out and stabilise open and Hoogsteen base pairs prior to cross-linking.

1.7: Oligopeptide antibiotics - netropsin, distamycin and the lexitropsins

The lexitropsins (also termed the 'hair-pin polyamides') are a class of non-covalent minor groove binders that have the ability to sequence selectively 'read' DNA. The first two examples were the di and tri-peptides **distamycin A** and **netropsin (Figure 29).**

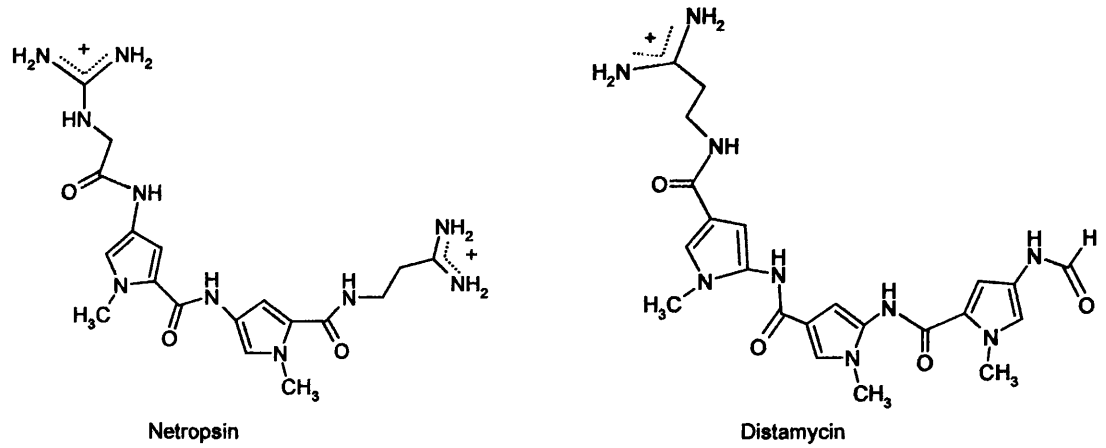


Figure 29: Chemical structures of netropsin and distamycin

Distamycin A and netropsin are naturally occurring peptide antibiotics isolated from *Streptomyces distallicus* and *Streptomyces netropsis* respectively.[143,144] Their structure comprises aromatic rings linked together by amide groups and both drugs display a wide range of antiviral, antitumour and antibacterial activities.[145,146]

Biological activity is exerted by tight binding to the minor groove of B-form DNA, causing interference with replication and transcription.[147,146] The drug molecules show a high preference for AT rich sequences, generally binding to clusters of more than four AT or IC base pairs,[148] with distamycin A binding preferentially to 5'-AAATT-3'[149,150] or 5'-AATT-3'.[151,152] Several studies including footprinting,[153,154] calorimetric,[155] crystallographic[156-158] and NMR[159-161] have revealed that netropsin preferentially binds to a 5'-ATTT-3' site, even when other sites with four AT base pairs are available.

Base specificity is thought to have its origins in *Van der Waals* forces between adenine C2 hydrogen atoms and the CH groups on the pyrrole rings of the drug molecule.[156] The drug molecule exhibits a small amount of torsional twist; this maximises contacts with the minor groove wall, with specific hydrogen bonds to the base edges providing AT specificity (**Figure 30**).[162]

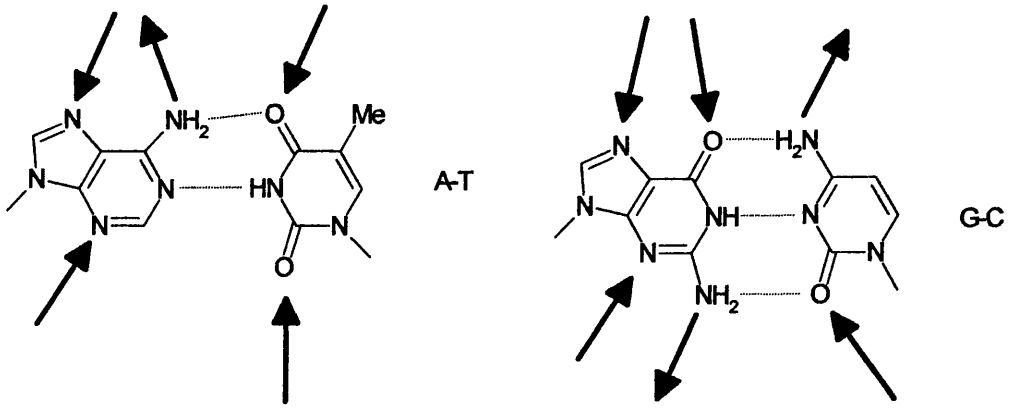


Figure 30: Specific hydrogen bonds from lexitropsin compounds to base edges providing AT specificity[162]

There is an additional electrostatic component from the cationic ends, but this is not deemed essential[146,163,164] and calorimetric studies with 5'*d*(CGCAAATTTGCG)$_2$[165] indicate that the process of binding is enthalpy driven with electrostatic interactions not dominant. Neither netropsin nor distamycin will attach in GC rich areas due to steric hindrance of the guanine NH_2 group.[147,166,167] Clashes are thought to occur between the NH_2 and H atoms of the pyrrole rings.[158] X-ray analysis has shown that the drugs bind within the minor groove by displacing the water molecules of the spine of hydration.[156,158] Binding occurs through a combination of interactions including hydrogen bonding, ionic charge attractions and *Van der Waals* forces.[158] The drug pyrrole rings are hydrophobic and increasing the number of N-methyl-pyrrole-carboxamide units appears to stabilize the drug-DNA complex.[163] Stacking interactions have been observed between the sugar O1' atoms and the three drug pyrrole rings and it is thought that this also contributes to the stability of the complex.[152]

Although the two naturally occurring oligopeptide antibiotics netropsin and distamycin display a marked preference for AT rich regions, it was recognised early on that strategic modification of the drugs could render them susceptible to the GC sites also. A large number of lexitropsins were synthesised and evaluated in which the pyrrole ring was replaced with N-methylimidazole.[168] This resulted in some recognition of GC sites but a lack of specificity between these and the AT rich regions, along with reduced DNA affinity probably related to the wider GC regions of the minor groove leading to weaker *Van der Waals* interactions.[169]

Several binding modes have been observed for distamycin and netropsin. These include binding in a 1:1, a 2:1 and even a 4:1 drug-DNA ratio.[171] In 1989 Pelton and Wemmer[170] used NMR to examine the binding of distamycin A to 5'*d*(CGCAAATTGGC).5'*d*(GCCAATTTGCG) and observed a concentration dependent change in binding modes. At low drug-DNA ratios the 1:1 complex dominated, but when drug concentration was increased a 2:1 binding mode was observed, with a drug binding affinity as high as that for the 1:1 complex. In a simple 1:1 binding mode three of the drug amide groups form a bifurcated (or three-centre) hydrogen bond between the adjacent adenine N3 or thymine O2 atoms in the A-T base pairs at the floor of the minor groove (**Figure 31**). Due to the narrowness of the AT rich

minor groove regions the drug sits symmetrically. Its two rings are slightly co-planar so that each ring is parallel to the walls of the groove.[152]

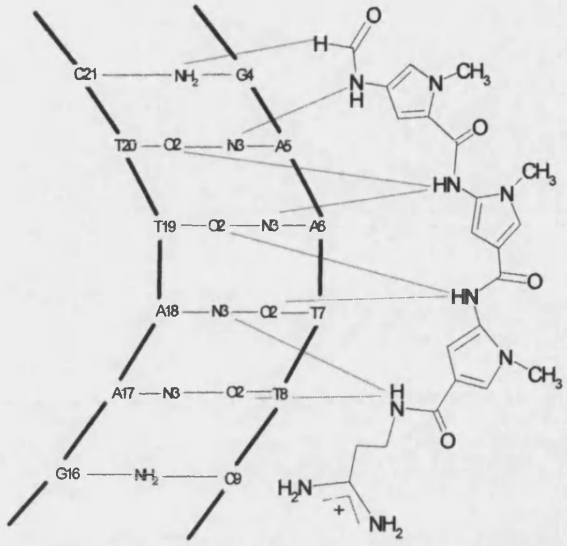


Figure 31: Bifurcating bonds in a distamycin-DNA complex[152]

The binding of lexitropsins to DNA has been observed to force open the minor groove by up to 17 Å.[171] In a 2:1 binding mode the drug molecules bind side-by-side, expanding the minor groove to at least 6.8 Å **(See Figure 32).** An anti-parallel orientation is favoured as it reduces charge-charge repulsion between positively charged propylamidinium groups.[170] Modeling suggests that the distamycin A molecules are staggered with respect to one another so that drug-DNA and drug-drug stacking interactions can contribute to the complex stability. Molecular modeling has also shown that the charged end of the drug is forced deeply into the minor groove where the electrostatic potential is larger[174] and a favourable contribution to the free energy of complex formation can therefore be made. As with the 1:1 complex hydrogen bonds between the drug amide and adenine N3 and thymine O2 also stabilize the complex, but because in the 2:1 binding mode the drugs are pushed to the walls of the duplex the bifurcated bonds are not observed.[170]

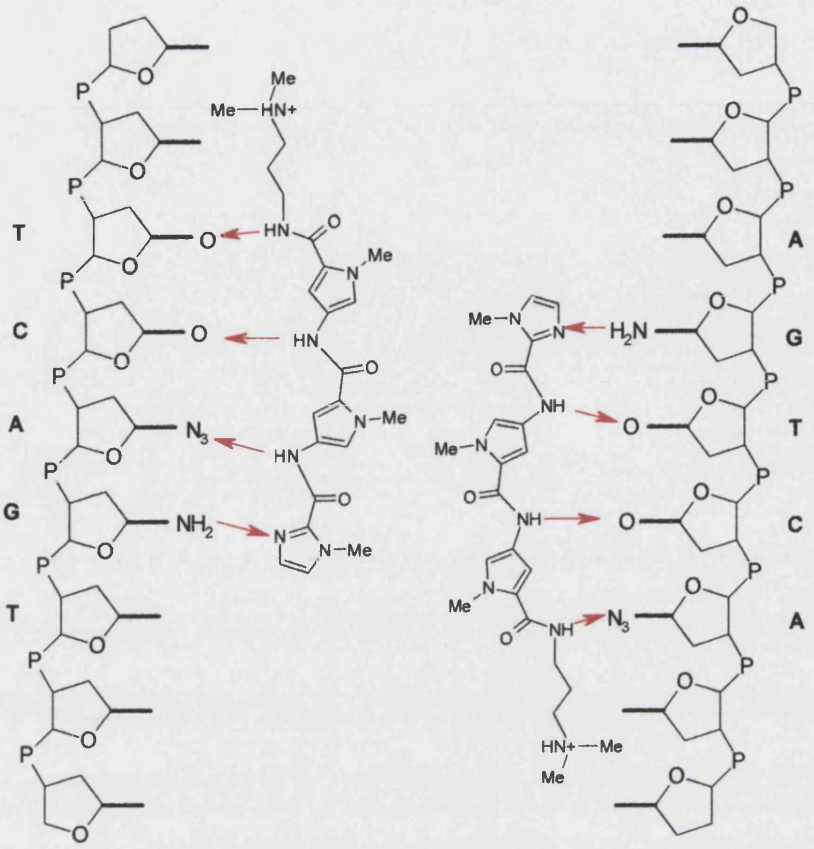


Figure 32: Schematic view of a side by side dimer, reproduced from reference 162

The 4:1 binding mode has been investigated using NMR and molecular modeling studies.[171] The DNA duplex accommodates four distamycin molecules bound in a tandem anti-parallel manner, existing in solution with free DNA and a 2:1 complex. No 3:1 or 1:1 complexes are seen in the mixture. The minor groove widens by 16-17 Å in order to accommodate the stacked drug molecules, an observation supported by extensive exchange of the A-T imino protons with the solvent at the binding site. The widening of the minor groove is accompanied by a change of only ± 2 Å in the major groove.

1.8: PBD-Lexitropsin conjugates

Recent work in the field of the lexitropsins and the PBDs has seen the synthesis and evaluation of a number of conjugate molecules of the two classes, displaying the PBD subunit linked to a number of combinations of pyrroles and imidazoles. In 2005 Kamal *et al* synthesised a series of PBD conjugates with a view to increasing water solubility and combining the effects of both the covalent (PBD) and the non-covalent (azepane) moieties (**Figure 33A**).[175] Varying linker lengths were evaluated, with one compound showing a significantly increased melting temperature when bound to DNA, as opposed to only marginal changes for other compounds. This once again highlights the importance of linker length in linked compounds designed for the binding of DNA.

Other conjugate molecules include nicotinamide-PBD hybrids (**Figure 33B**),[176] all showing increased DNA binding relative to DC-81, a series of PBD hybrids linked with indole carboxylates (**Figure 33C**)[177]– again all showing potent antitumour activity, with hybrids possessing a greater toxicity than DC-81. It is thought that the unity of the two drug classes increases recognition for DNA binding sites and hence increased complex stability is observed.

C8-linked PBD-poly(N-methyl pyrrole) conjugates have also been reported (**Figure 33D**).[178] These compounds show an increase in DNA binding affinity that is more than 50-fold greater than the individual PBD and pyrrole components. All of the conjugates studied appeared to take on the properties of both component fragments, showing sequence-selective DNA binding with high affinity, as well as an ability to block RNA transcription.

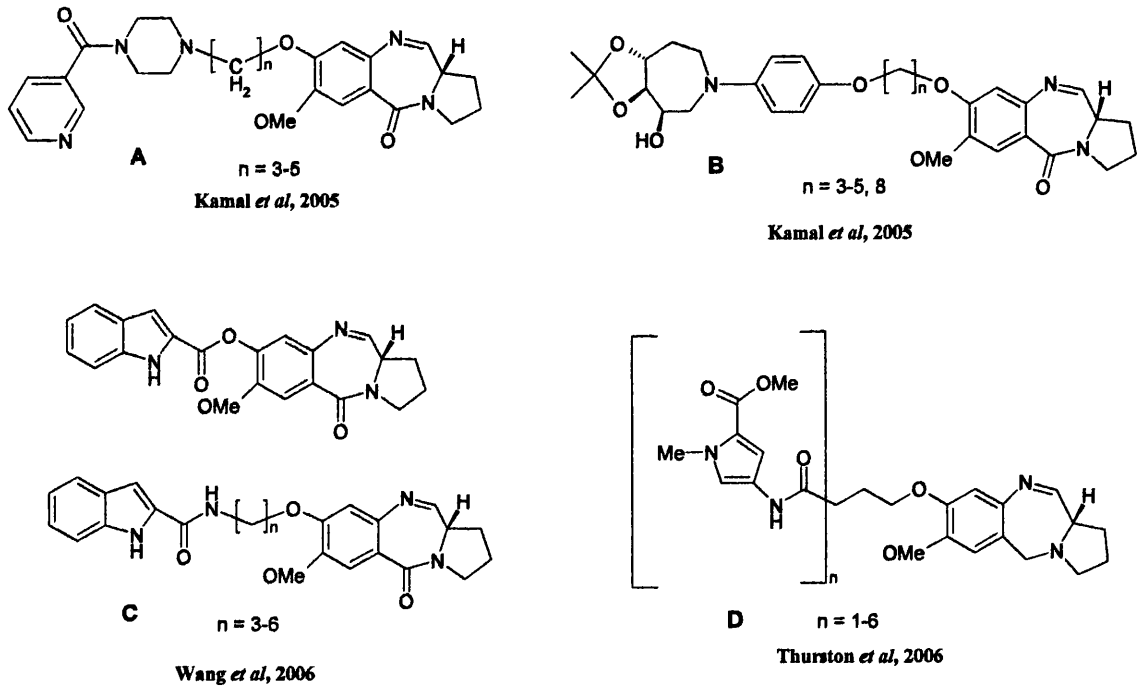


Figure 33: PBD conjugates synthesized in recent years[175-178]

1.9: Aim of this project

1.9.1: Synthesis and analysis of drug-DNA adducts with SJG-136

As discussed previously, synthetic cross-linking PBD dimers such as DSB-120 and SJG-136 are designed to be inter-strand cross-linkers. However, work carried out by Cameron and Thompson[179,180] suggested the possibility of an intra-strand cross-link. This work involved the reaction of DSB-120 with a 5'*d*(CTCATCAC).(GTGATGAG) sequence, chosen due to the possibility of an intra-strand cross-link site utilising covalent modification of G11 and G14 on the same strand.

This work was not taken to completion due to the degradation of the PBD dimer, but on reaction with the drug a broad hump appeared at 8.70 ppm in the ^{1}H NMR spectrum of the drug-DNA adduct (**See Figure 34**).

Mountzouris *et al*[181] had noted a similar peak in 90 % D_2O at 9.10 ppm in a 5'*d*(CICGATCICG)$_2$-DSB-120 inter-strand cross-linked adduct. This was thought to be due to the NH proton of guanine N2, which is covalently bound to DSB-120 at DSB-C11 *via* an aminal linkage. The presence of this peak in the spectrum produced by Cameron and Thompson supported the covalent modification of the duplex by DSB-120, though it was not possible on the strength of the data to determine whether the drug had bound at both G11 and G14, or just at one.

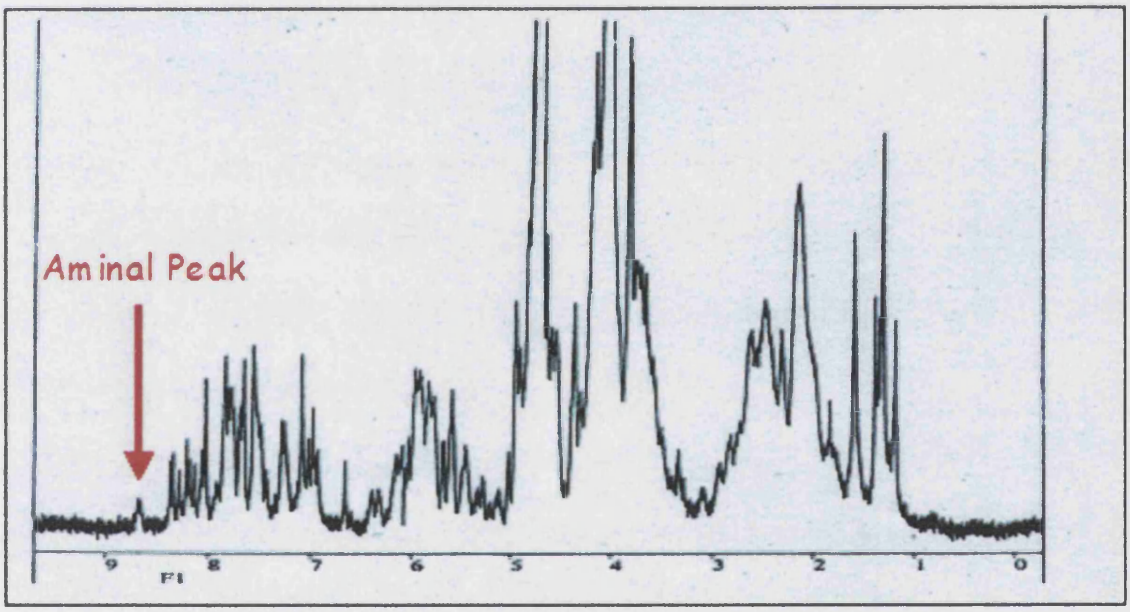


Figure 34: ^{1}H NMR of DSB-120-5'*d*(CTCATCAC).5'*d*(GTGATGAG) adduct (Cameron and Thompson)[179]

It was proposed during this project to synthesize the 5'*d*(CTCATCAC).(GTGATGAG) duplex for reaction with SJG-136, in order to investigate further the possibility of the intra-strand cross-link and to compare the drug-DNA adduct with that of the inter-strand linked DNA formed with DSB-120.

1.9.2: Synthesis and analysis of further drug-DNA adducts with adozelesin

Previous investigation of a 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct by high field NMR saw evidence suggestive of an overlap of the ends of the two adozelesin molecules in the *bis*-adduct.[179,180] Molecular modeling studies as a result of NMR data

were performed on one half of the symmetrical *bis*-adduct, and showed slight distortion at the ends of the molecule, with the benzofuran 'C' subunit exhibiting a slight 'fish-tailing' effect. It is possible that this 'fish-tailing' could be the result of a stacking interaction between the aromatic rings of the benzofuran subunit, as shown in **Figure 35.**

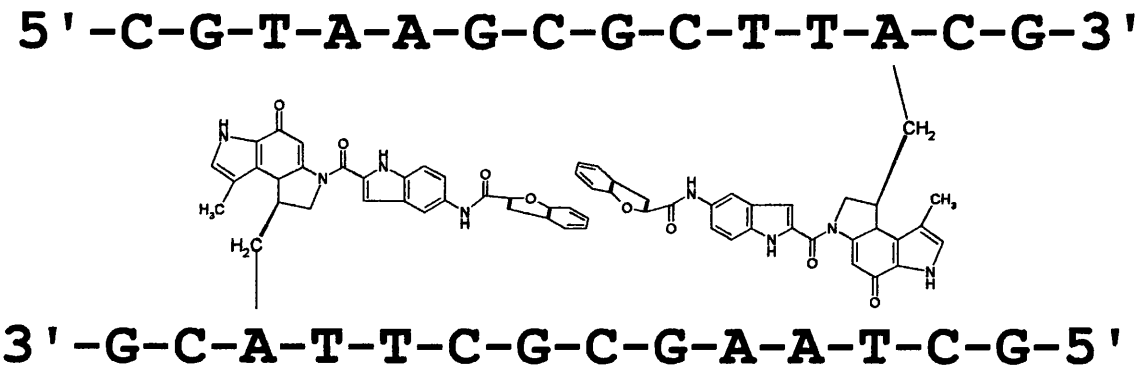


Figure 35: Suggested overlap of benzofuran subunits in the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct

Incidences of molecular 'stacking' have been discussed previously in conjunction with the lexitropsins, and a comparison of the adozelesin molecule with this class of drug yields several similarities (**See Figure 36**).

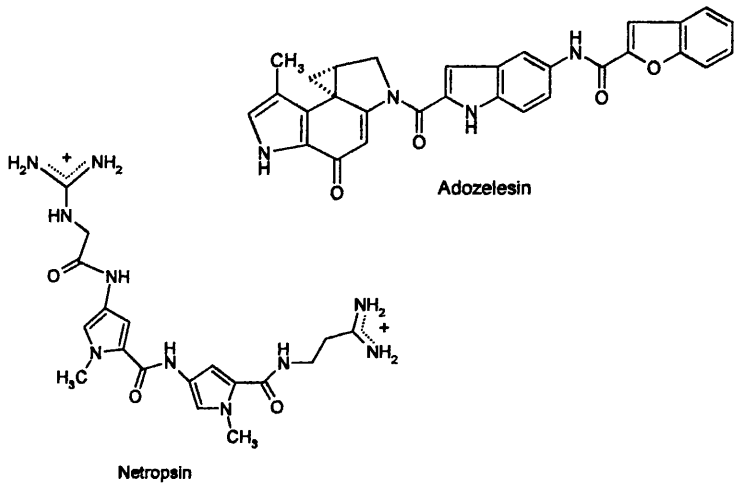


Figure 36: Comparison of the molecular structures of netropsin and adozelesin

It can be seen that both drugs contain planar units linked by flexible linkers, which allow the drugs to bend with the turn of the DNA helix. As stated previously, lexitropsins are accommodated in the minor groove by the utilisation of hydrogen bonding and *Van der Waals* forces. Conversely adozelesin forms a covalent attachment, although hydrogen-bonding interactions are also noted between the 'B' and 'C' subunits and the DNA duplex. Due to the shape of the drug and the fact that it is held in place by a covalent bond it is suggested that it may be possible for this molecule to at least partially 'stack' in a similar manner to the lexitropsins.

It was proposed during this project that a shorter DNA duplex would be synthesized in the form of 5'*d*(CGATTAATCG)$_2$. This duplex is 4 base pairs shorter than that previously used to bind adozelesin, and so in theory could produce significant overlap of the drug molecules.

CHAPTER 2 – EXPERIMENTAL PROCEDURES

2.1: Laboratory reagents – sources

β-Cyanoethyl phosphoramidites (deoxy: adenine, thymine, guanine, cytosine and inosine ie dA, dT, dG, dC and dI) – Cruachem
Iodine (oxidising agent) – Cruachem
Trichloroacetic acid (TCA), (deblok) – Cruachem
Acetic Anhydride (Cap A) – Cruachem
4-(Dimethylamino) pyridine (DMAP), (Cap B) – Cruachem
Tetrazole (Activator) – Cruachem
Aqueous ammonia (NH_4OH) – Sigma 35% NH_3
Ammonium acetate (NH_4OAc)- Sigma anal. grade 99%
Acetic acid (CH_3COOH) – Fisons HPLC grade
Acetonitrile (CH_3CN) – Fisons HPLC grade
Controlled pore glass (CPG) support – Cruachem
Deuterium oxide (D_2O) - Sigma
Disodium hydrogen orthophosphate ($Na_2HPO_4.2H_2O$) – Fisons anal. grade, 98 – 100%
Sodium dihydrogen orthophosphate ($NaH_2PO_4.2H_2O$) – Fisons anal. grade, 98 – 100%
Sodium Chloride (NaCl) – Sigma, 99.9%
Methanol (CH_3OH) – Fisons HPLC grade
Dimethylformamide-d_7 – Sigma
Adozelesin – A gift from the former Upjohn Company, Kalamazoo, Michigan 49001, USA
SJG-136 – A gift from **IPSEN**, France

2.2: Synthesis of DNA

2.2.1: Preparation

β-Cyanoethyl phosphoramidite bases were dissolved in HPLC grade acetonitrile (ACN) as shown in **Table 2.**

Base	Weight (g)	Vol of ACN (mL)	Molarity
Adenine	1	11.2	0.1
Guanine	1	11.6	0.1
Cytosine	1	11.8	0.1
Thymine	1	13.2	0.1

Table 2: Dissolving of bases in acetonitrile

2.2.2 Synthesis

The required DNA monomer (15μm) was synthesised using phosphoramidite chemistry on an *Applied Biosystems 381A DNA Synthesiser.* The synthesiser was programmed to synthesise "trityl-on" DNA to enable subsequent purification of the sample using HPLC. The deprotection step was monitored by UV spectroscopy (λ = 497) in order to check that the coupling efficiency was satisfactory. The four stages of oligonucleotide synthesis are shown in **Figures 37-40:**[182]

A. Detritylation **(Figure 37)**

In this step the dimethoxytrityl protecting group is removed from the 5'-OH of the nucleoside anchored to the column. The deprotecting agent used is trichloroacetic acid (TCA) (obtained from *Cruachem* under the name "**Deblok**").

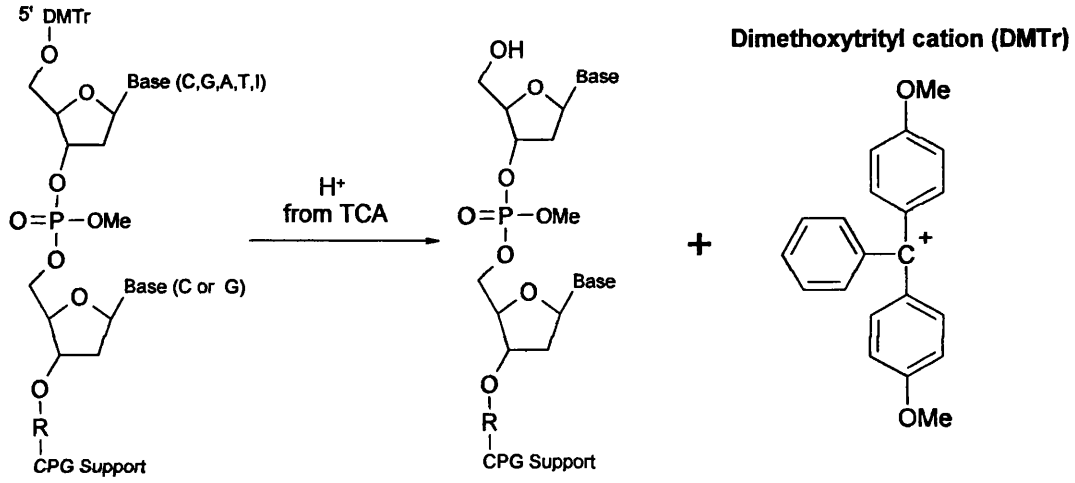


Figure 37: Removal of dimethoxytrityl group

B. *Addition* **(Figure 38)**

In the addition step a 5' to 3' inter-nucleotide linkage is formed between the first and second nucleotide in the sequence *via* a trivalent phosphorous. Tetrazole is added to the column simultaneously with the nucleoside derivative. The tetrazole protonates the nitrogen of the diisopropylamine group on the 3' phosphorus, rendering it susceptible to nucleophilic attack by the detritylated 5' –OH of the strand attached to the support.

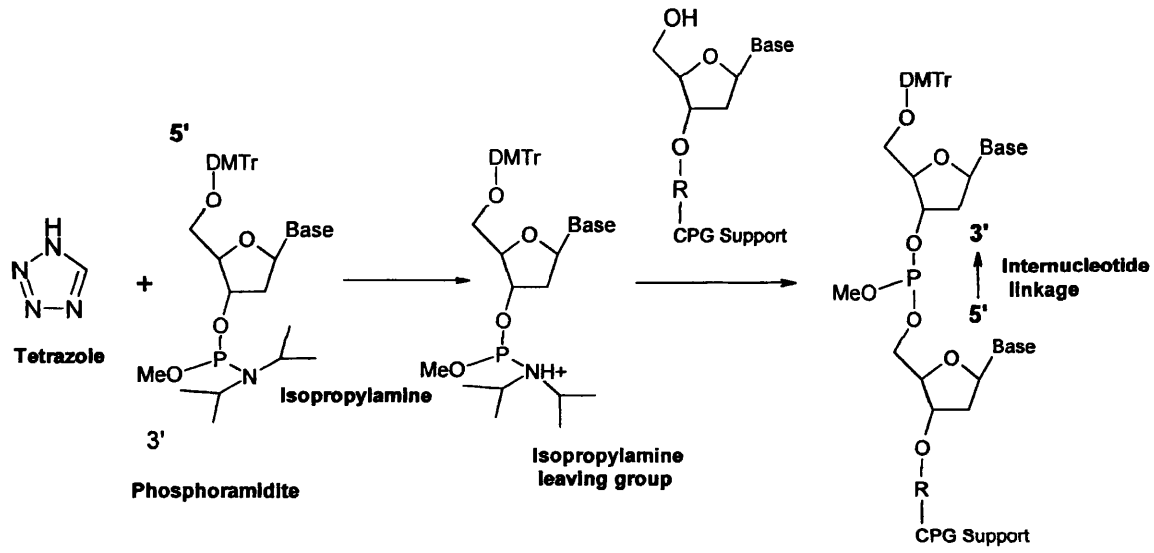


Figure 38: Addition

C. *Capping* **(Figure 39)**

The coupling reaction has an efficiency of 98 - 100 %. However, some chains will not undergo addition and a capping step is necessary to terminate these in order to limit the lengths of any impurities and enable separation after synthesis. Unreacted chains will all have a free 5'-OH, whereas in the reacted chains this will be blocked with a dimethoxytrityl group. The free -OH groups are capped using an acetylating agent

formed from equal volumes of equimolar acetic anhydride and dimethylaminopyridine (DMAP).

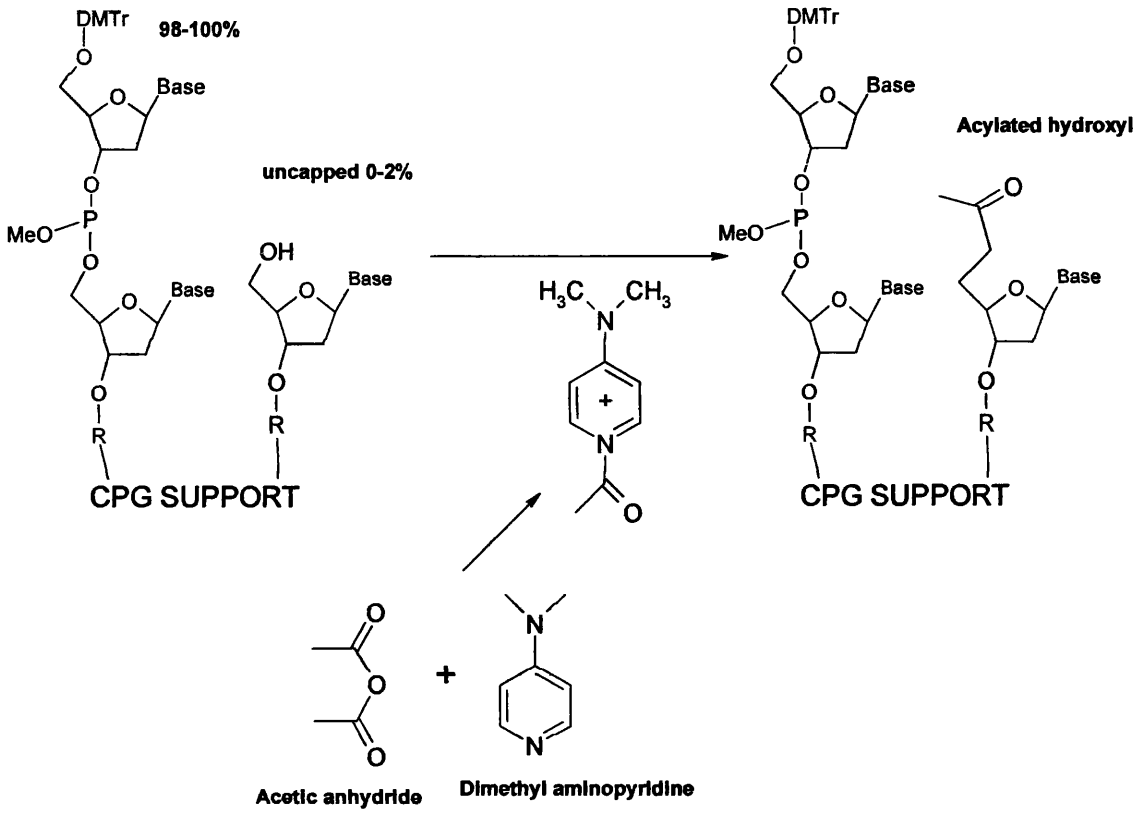


Figure 39: Capping reaction

D. *Oxidation* (**Figure 40**)

The trivalent phosphorus (phosphite) triester linkage formed in the coupling reaction is unstable and susceptible to acid and base cleavage. To combat this problem conversion to a more stable pentavalent phosphate triester is performed immediately after the capping step. This also helps to avoid the formation of acetic acid around the oligonucleotides (caused by water from the oxidising solution mixing with acetic anhydride). The oxidising solution is a basic lutidine/THF solution, in which iodine acts as the oxidising agent and water the oxygen donor. An iodine lutidine complex is formed with the trivalent phosphorus, this is then displaced by water to form a pentavalent phosphorus. The whole reaction takes less than 30 seconds.

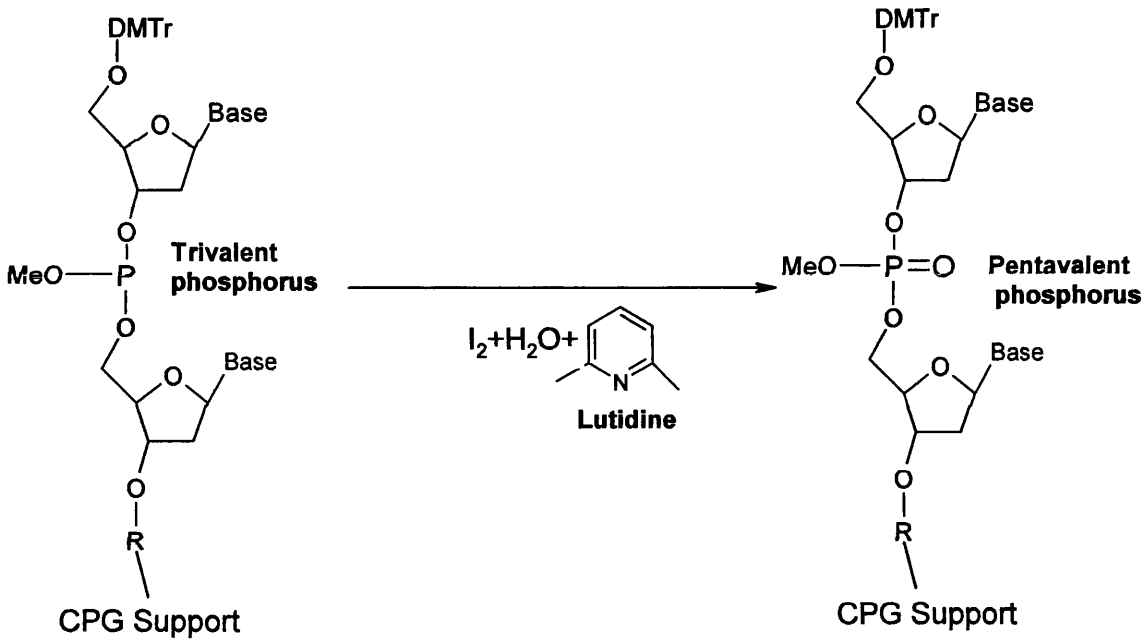


Figure 40: Oxidation

These four steps are repeated for each nucleoside addition. The end result is an oligonucleotide strand attached to the CPG column at the 3' terminal and protected by a dimethoxytrityl at the 5' terminal. This strand is now removed from the CPG support and then requires purification before it can be used for drug addition.

2.3: Stages in the Purification and Annealing of DNA

2.3.1. Removal of DNA from the CPG Support

The DNA sample was removed from the CPG support following instructions given in the *Biosystems* manual.[182] Two syringes were connected to either end of the column and 5mL of concentrated ammonia (35 %) was flushed back and forth across the column several times to ensure that the packing was thoroughly wet. The column was then left with the syringes depressed half-way for half an hour. This was repeated twice, after

which the column was repeatedly flushed with further aliquots of ammonia. The total ammonia used was around 50 mL.

During this step the DNA sample is cleaved from the column and the β-cyanoethyl protecting groups are removed (**Figure 41**). This step was monitored by UV spectroscopy (λ = 254) in order to ascertain the point when all DNA had been removed from the column.

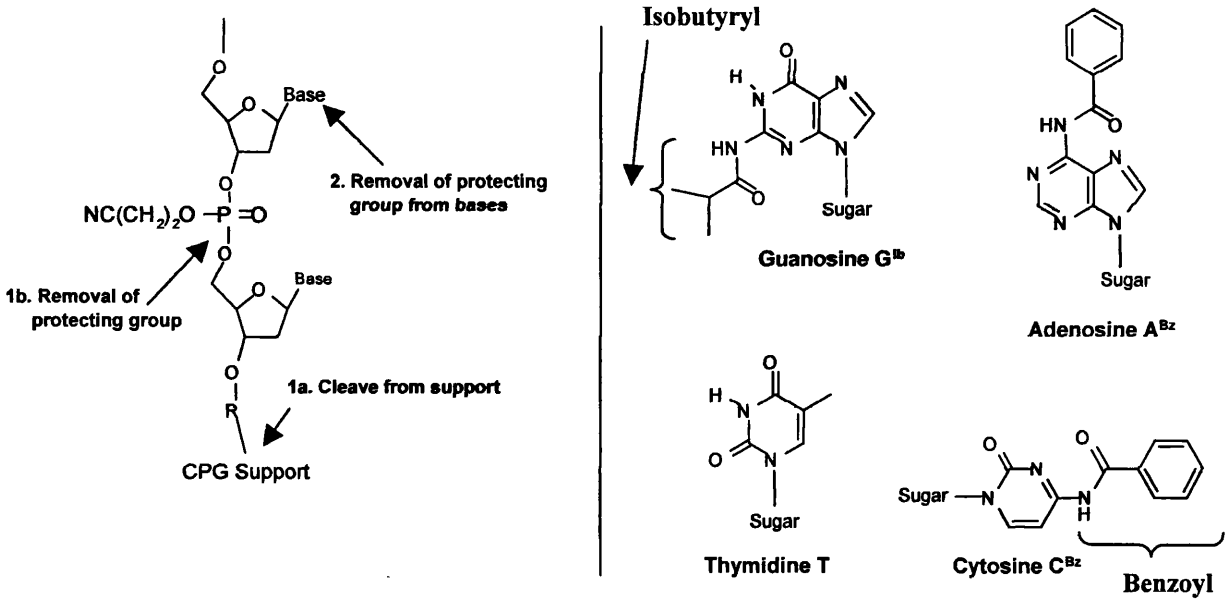


Figure 41: Removal of DNA from support and deprotection

2.3.2: Deprotection of the Bases

Bases are protected using isobutyryl and benzoyl groups as shown in **Figure 41.** Deprotection of bases was achieved by heating of the DNA sample in concentrated ammonia for 16 hours at 55°C. This was done overnight, by heating the ammonia sample in a sealed, Schott bottle in a water-bath thermostatted at 55 °C.

2.3.3: Purification by HPLC

Purification of the single-stranded DNA was carried out using reverse phase HPLC with a Dynamax-300A, C18 (20mm x 200m) preparative HPLC column. The failed syntheses were separated from the pure DNA strand using an HPLC linear gradient system from 0.1M ammonium acetate to 40 % *v/v* acetonitrile (HPLC grade) in 0.1M ammonium acetate over a period of 90 minutes. HPLC conditions are given in **Table 3.**

GRADIENT (% *V/V*)	0-40
Total Time (mins)	90
Wavelength (nm)	254
Flowrate (mL/min)	10
Recorder sensitivity (dual recorder)	10mv (red), 2V (black)
Recorder speed mm/min	10

Table 3: HPLC conditions for DNA purification

HPLC of the tritylated sample produced a characteristic trace, independent of the sequence of bases. (**See Figure 42**).

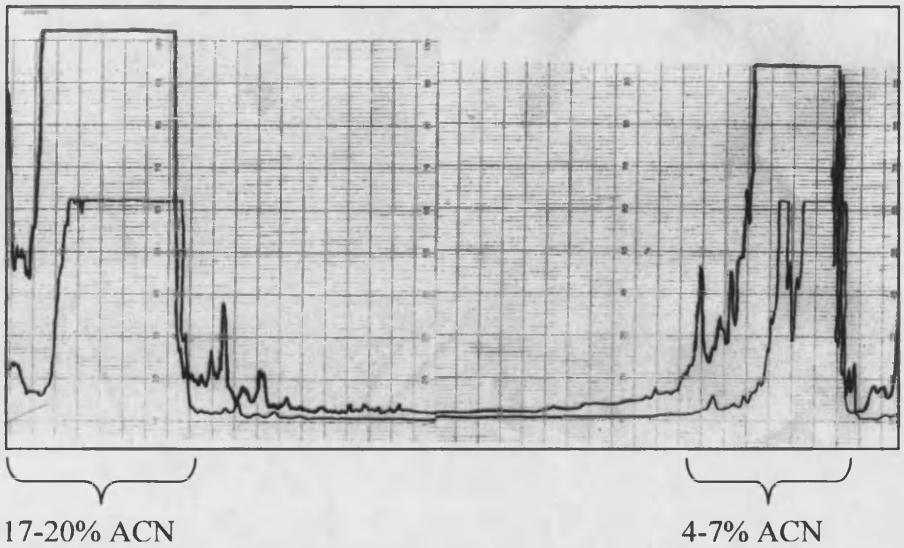


Figure 42: Characteristic HPLC trace showing separation between failed syntheses and pure DNA. Right hand peaks = eluted first.

The failed syntheses contain no trityl group, and consequently elute at between 4% and 7% acetonitrile as a series of sharp off-scale peaks, corresponding to capped strands of increasing molecular weight. The pure strand of DNA is eluted last at about 17 % – 20 % acetonitrile. There is a large gap between failed and complete synthesis connected to the polarity of the molecules. The capping reaction results in failed syntheses having an acetylated 5'-OH instead of a dimethoxytrityl group (**See Figure 43**). The smaller acetyl group is more polar than the dimethoxytrityl group and thus acetylated strands are partitioned into the acetonitrile mobile phase at a lower concentration.

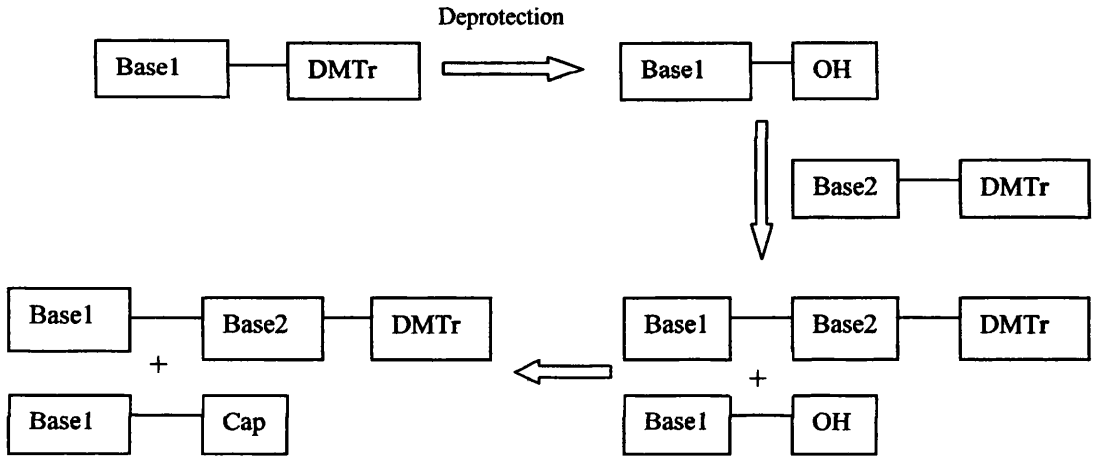


Figure 43: Diagram showing capping procedure – resulting in only pure DNA strand having DMTr group

The DNA containing fraction was collected and concentrated to around 5 mL using a rotary evaporator. The sample was then dialysed against 1.5 L Milli Q water for 16 hours using dialysis membrane (MWCO = 1000) in order to remove the salt. The water was once again removed by rotary evaporation.

2.3.4: Detritylation

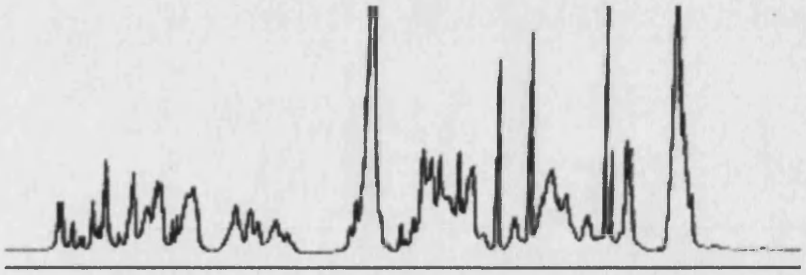
The DNA strand was detritylated by treatment with 50 % *v/v* glacial acetic acid (dilute in Milli Q) for 20 minutes. A transient pink colour was observed. Acid was then

removed by rotary evaporation and the sample dialysed in 1.5 L MilliQ water, in order to remove the cleaved trityl groups. The quality of the sample was then checked using 1D proton NMR.

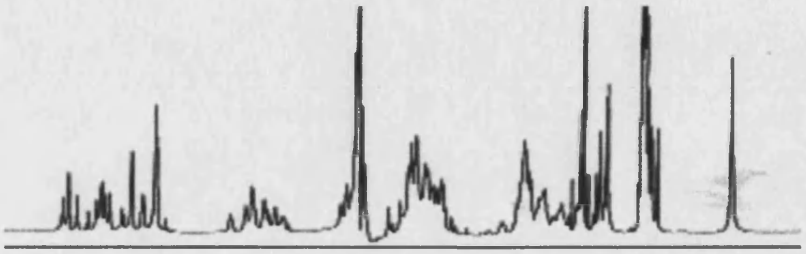
If at this point further purification was required a second HPLC step was then performed as before, but with a gradient of 0-10 % *v/v* acetonitrile in ammonium acetate over a period of one hour. The collected DNA peak eluted at around 5 % acetonitrile and was concentrated and dialysed as before.

2.3.5: Examination by NMR

Prior to annealing it was necessary to add 0.6 mL of NMR buffer to each oligonucleotide strand. This is important as salt concentration has a significant effect on spectral resolution, and also helps to sustain the three-dimensional duplex structure of the DNA, thus avoiding formation of hairpins at low ionic strengths.[183] NMR buffer was prepared by mixing 0.01M NaH_2PO_4 and 0.01M Na_2HPO_4 (made up in 0.1M NaCl) until they reached pH 6.85. This phosphate buffer will maintain pH, but its own protons are likely to be responsible for additional peaks in the NMR spectrum. This problem was addressed using exchange of these protons by redissolution in D_2O. The DNA sample was first dissolved in 0.6mL NMR buffer, then evaporated to dryness, dissolved in 1 mL D_2O, freeze dried again, and finally taken up in 0.6 mL D_2O. ^1H NMR was used to examine the samples, and if clean enough they were ready to be annealed into a duplex. Examples of satisfactory ^1H proton NMR of a DNA strand before and after purification are shown in **Figure 44**.



Unpurified 5'*d*(GTGATGAG) on removal from CPG support



5'*d*(GTGATGAG) after HPLC, dialysis and detritylation steps

Figure 44: Comparison of DNA duplex strands before and after purification by HPLC

2.3.6: Annealing

For non-palindromic sequences such as 5'*d*(CTCATCAC).(GTGATGAG) it was necessary to anneal the samples by mixing equal quantities of each complementary strand. Quantities of DNA were determined using absolute signal to noise measurement in the NMR spectrum. Each strand was dissolved in 0.3 mL 100 atom % D_2O and the two were mixed together in an NMR tube. The mixed strands were then heated to around 70°C by immersion of the NMR reaction tube in hot water in a flask and allowed to cool over 3-5 hours. As the samples are heated the DNA melts (midpoint melting temperature is *ca.* 69°C[16]) and reforms as duplex DNA as the sample cools down. For palindromic sequences such as 5'*d*$(CGATTAATCG)_2$ the process involves simply dissolving the strand in 0.6 mL 100 atom % D_2O and heating as before.

2.4: DNA adduct formation

On production of a pure DNA duplex the DNA was treated with a drug in order to produce a covalently bound drug-DNA adduct.

2.4.1: Reaction of 5'd(CTCATCAC).(GTGATGAG) and 5'd(CICGATCICG)$_2$ with SJG-136

The method for drug addition to DNA outlined by Mountzouris *et al*[181] was followed:

The duplex in question was freeze-dried from double-strength NMR buffer (as outlined in section 2.3.5). It was then taken up in 0.6 mL D_2O, with the addition of 0.2 mL deuterated methanol (100% atom) to aid solubility of the drug. The duplex was then added to 5 mg of SJG-136 – constituting a large drug excess. The mixture was stirred for 16 hours at room temperature in the absence of light.

On reaction of the drug (confirmed by ^1H NMR) the reaction mixture was filtered and excess drug removed. The resulting adduct was then subjected to further NMR experiments as outlined in section 2.4.3.

2.4.2: Reaction of 5'd(CGATTAATCG)$_2$ with Adozelesin

The duplex was freeze-dried from double-strength NMR buffer as outlined in section 2.3.5 and taken up in 0.6 mL D_2O. The duplex was added to 4 mg of adozelesin dissolved in 0.4 mL deuterated DMF (constituting a large excess of drug). The mixture was initially stirred for 48 hours in the absence of light. On assessment of the ^1H NMR spectrum it was ascertained that the reaction had not proceeded to completion and so the mixture was allowed to continue reacting for a further 10 days. At this point ^1H NMR was utilised to confirm completion of the reaction and the excess drug was filtered from the mixture. The resulting adduct was subjected to further NMR experiments as detailed in section 2.4.3.

2.4.3: NMR Conditions

All NMR experiments were carried out in double-strength NMR buffer as detailed in section 2.2.5. One and two-dimensional 400 and 600 MHz NMR data sets were recorded in H_2O and D_2O on Varian 400 and 600 spectrometers. Proton signals were

recorded in parts per million (ppm) and referenced relative to the residual water signal at 4.71 ppm.

Phase-sensitive NOESY spectra (600 MHz) were obtained for mixing times 200 and 300 ms. Spectra were recorded with 32 scans at each of 1024 t_1 values at a spectral width of 10.002 ppm at a relaxation delay of 2 s between scans. Two-dimensional NOE spectra in 90 % H_2O at 150 ms mixing time were recorded using water suppression by pre-saturation with a 2 s repetition delay and a sweep width of 25.025 ppm.

DQF-COSY spectra were also recorded on the Varian 600 MHz spectrometer and were used to confirm assignments made in the NOESY spectra. The 5'*d*(CGATTAATCG)$_2$-adozelesin adduct was also subjected to ROESY experiments in order to rule out any conformational exchange within the duplex.

2.5: NMR Experiments

2.5.1: Basic NMR Spectroscopy[184-186]

The first successful NMR experiment was demonstrated in 1946, and since then the technique has become an indispensable research tool. The use of NMR for studies of drug-DNA interactions carries a distinct advantage in that samples are in solution. This means that important parameters such as pH, temperature and ionic strength can be varied in order to mimic the conditions experienced in a biological system.

NMR spectroscopy is based on the fact that nuclei that contain an odd number of protons or neutrons possess a magnetic moment. When an external magnetic field is applied the magnetic moment becomes aligned either parallel or opposed to the direction of the applied field. In the case of the hydrogen atom, ^1H, this corresponds to the two spin states + ½ and – ½ . The two spin states have different energies, and an exact amount of energy is required to 'flip' the magnet from one position to the other. This amount of energy, ΔE, is related to the magnetic moment of the proton, μ, the applied field strength, B_0 and the frequency, ν, as shown in **Equation 1.**

$$\Delta E = h\nu = 2\mu B_0$$

Equation 1: Equation relating amount of energy ΔE to change the spin state of a proton with magnetic moment μ, applied field strength B_0 and frequency ν.

Early NMR experiments were carried out by keeping a constant frequency and varying the magnetic field in order to determine the point where the magnetic moment flipped. Modern spectrometers however, work by irradiating the sample with a short pulse of electromagnetic radiation, hence delivering a range of frequencies simultaneously to the sample while keeping the applied magnetic field constant. The ^1H nuclei are excited, and as they relax back to equilibrium conditions a signal relating to each excited frequency can be detected. A technique known as Fourier Transform is then employed to convert the signals detected into the final spectrum, where protons appear as resonances at different chemical shifts according to their environment.

2.5.2: The Nuclear Overhauser Effect

The Nuclear Overhauser Effect (NOE) is due to a transfer of magnetisation between two spins that are in close spatial proximity. If two spins, A and B, are within 5 Å of each other magnetisation is transferred between them by through-space cross relaxation. This means that irradiation of one nucleus at its resonance frequency affects the strength of the signal arising from the second nucleus – it is either strengthened (Nuclear Overhauser Enhancement) or weakened. This means that it is possible to gain a wealth of information on the spatial locations of protons within a molecule.

It is possible to illustrate the theory behind NOE using a molecule containing two inequivalent protons, I and S, that are not scalar coupled. **Figure 45 A** shows the energy level diagram for these protons, as well as their equilibrium population distribution. As shown, there are six relaxation pathways possible. Four of the relaxation pathways correspond to single spin flipping, and are denoted W_1^I and W_1^S, where 'W' denotes the probability of a single quantum transition, while the subscript gives the change in

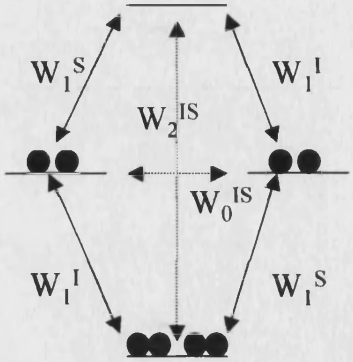
quantum number. The remaining relaxation pathways are due to cross-relaxation, when the I and S protons relax together. These are denoted W_0^{IS} and W_2^{IS}. If the sample is irradiated at the frequency of the 'S' proton the 'S' spin becomes saturated and the population distribution is that shown in **Figure 45 B**. It is now possible for cross-relaxation to occur *via* the W_2^{IS} and W_0^{IS} pathways. If relaxation occurs *via* W_2^{IS} the population of levels will be that shown in **Figure 45 C**, and will therefore result in an enhancement of the I signal. If the W_0^{IS} pathway is utilised then there is a decrease in signal as shown in **Figure 45 D**.

NOE experiments are not always an accurate measure of distance as there are problems that can arise. Low sensitivity (a low signal to noise ratio) can render integral calculations difficult, and problems can occur with larger molecules. A large molecule rotates slower than its smaller counterpart, and this can lead to spin-diffusion pathways (W_1^S and W_1^I) becoming as prominent as the cross-relaxation pathways. This changes peak intensities and complicates interpretation of distances. An increased mixing time can be used to make longer range NOEs easier to observe, but this will also make the spin-diffusion pathways more prominent in the shorter range couplings.

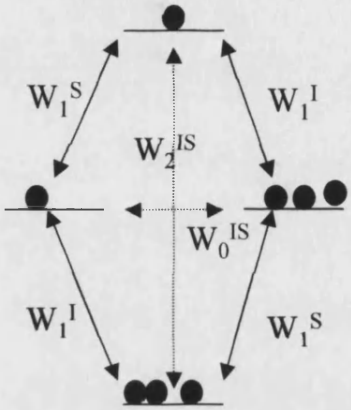
2.5.3: Two-dimensional Nuclear Magnetic Resonance

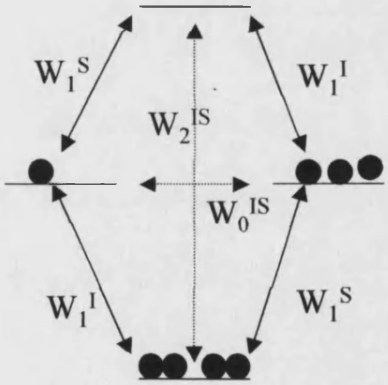
Conventional NMR spectroscopy results in spectra, which are a plot of intensity against frequency. For coupling nuclei however, such as H-H, the interactions are also a function of time. By sampling at a range of times therefore, it is possible to separate interactions in order to establish which nuclei are coupled. Two frequency axes are involved, and so the process is known as two-dimensional NMR spectroscopy. There are many different 2D-NMR experiments, but several are of particular use in the NMR of DNA, and will therefore be introduced.

A: Energy levels and equilibrium population distribution for a pair of spin ½ nuclei I and S

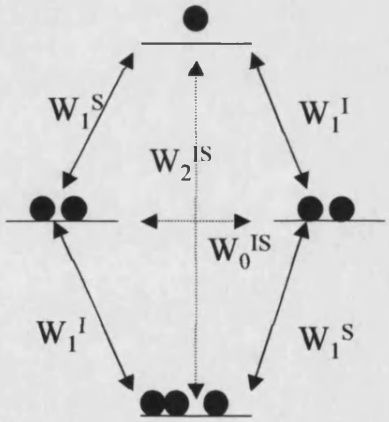


B: Population distribution after saturation of the 'S' spin





C: Relaxation *via* the W_2^{IS} pathway – enhancement of 'I' signal



D: Relaxation *via* the W_0^{IS} pathway- weakening of 'I' signal

Figure 45: Energy levels, relaxation pathways and population distributions of two dipolar coupled protons, I and S. Adapted from reference 184

2.5.4: Correlated Spectroscopy (COSY)

Through-bond coupling is known as scalar coupling; protons within three bonds of each other will exhibit $^{3}J_{HH}$ scalar coupling. The COSY experiment is used to collect information on protons that are scalar coupled. A spectrum is produced with the original 1D resonances pictured as a diagonal from corner to corner. Any scalar coupled protons will exhibit an off-diagonal cross-peak as shown in **Figure 46**.

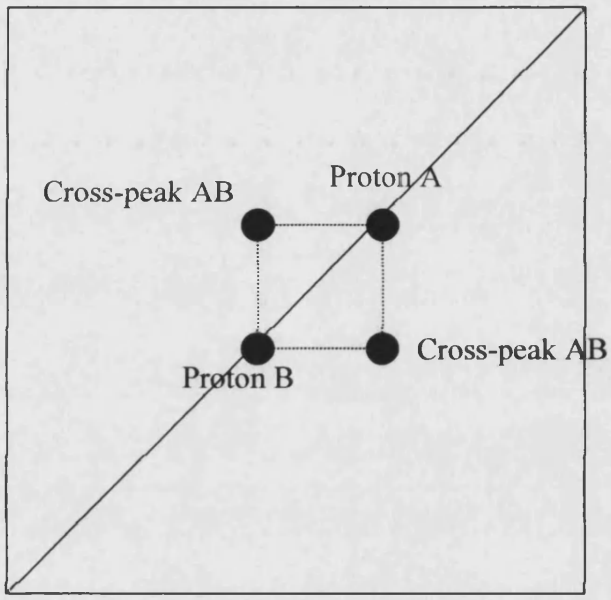


Figure 46: Cross-peaks shown by two scalar coupled protons, A and B, in a COSY experiment

The strength of the coupling observed is affected by several factors including the electronegativity of attached substituents and the dihedral angle. The effect of the dihedral angle is of particular importance in the NMR of DNA as this can cause the absence of expected COSY cross-peaks, making assignment of DNA by NMR more difficult.

The effect of the dihedral angle on the vicinal coupling constant can be effectively predicted using Karplus's equations (**Equation 2**).

The effect can be more readily visualized when the equation is expressed graphically as in **Figure 47**. It can be seen that the largest vicinal couplings arise when the dihedral angle is 0° or 180°. Protons at 90° to each other give rise to very small couplings and in this situation the expected COSY cross-peaks may be entirely absent.

$$J = A + B \cos \emptyset + C \cos 2\emptyset$$

Equation 2: Karplus equation[187] relating coupling constant, *J* to dihedral angle, ø

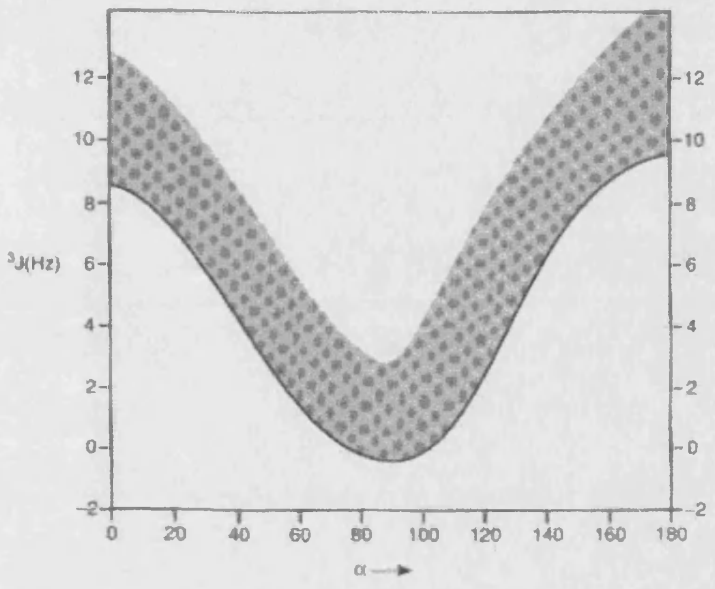


Figure 47: Graphical representation of Karplus equation[188]

2.5.5: Nuclear Overhauser Enhancement Spectroscopy (NOESY)

NOESY experiments are used to detect protons that are dipolar coupled, and are therefore useful in detecting protons that are close together in space. The NOE experiment works on the basis of the NOE theory discussed previously, and once again produces a two-dimensional spectrum in which the 1D spectrum is present as a central diagonal. Any protons within a spatial proximity of 5 Å will produce off-diagonal cross-

peaks similar to those seen in the COSY experiment. An estimate can be made as to the distance between the protons in question based on the intensity of the cross-peak observed.

2.5.6: Rotating Frame Overhauser Spectroscopy (ROESY)

ROESY is a similar experiment to NOESY – detecting through space couplings by use of NOE effects. However, in addition the ROESY spectrum provides information on conformational exchange. If exchange takes place during the NMR timescale peaks are visualised as opposite sign (ie positive peaks become negative). This is useful in detection of phenomena such as base ring opening or open duplex conformations in a DNA molecule.

2.6: NMR studies on DNA duplexes

There are several stages in the sequential assignment procedure for nucleic acids, which have been well documented.[189-192] For the purpose of discussion the 5'*d*(CTCATCAC).(GTGATGAG) duplex will be used for an in depth account of these stages.

2.6.1: The one-dimensional ^1H NMR spectrum

On collection of NMR data it is important to assign all non-exchangable protons before drug addition to the duplex, to enable a comparison to take place on addition of the drug. A diagram showing the numbering system for the duplex is shown in **Figure 48**.

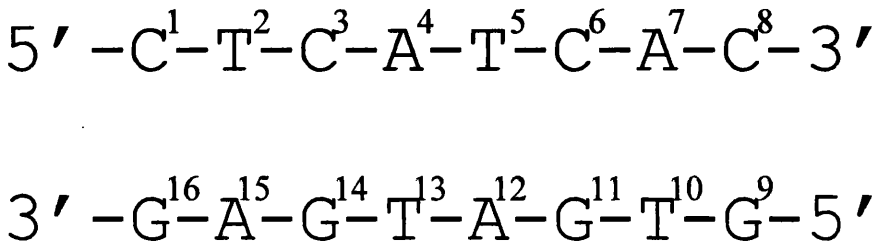


Figure 48: Diagram showing numbering system of 5'*d*(CTCATCAC).(GTGATGAG) duplex for NMR assignment purposes

The one dimensional (1D) spectrum is of limited use in the assignment of protons in DNA, due to the large incidence of overlaid peaks, making it impossible to distinguish one proton from another in much of the spectrum. The 1D spectrum therefore, is mainly used for quality control. An indication of sample purity can be obtained, with impure samples showing obvious impurities, line broadening and a low signal to noise ratio. These are all indications of the sample needing further purification or re-synthesis. It is also possible to determine whether correct nucleotide components are present, even if it is not possible to decipher their order. For this the base protons of each nucleotide are used, these appear in the aromatic region and are usually fairly well defined. Peaks due to protons of the DNA sugar backbone however cover two thirds of the spectrum and are greatly overlaid. It is not possible to do more than relate each spectral region to its respective deoxyribose proton in these regions (**See Figure 49**).

The Pu-H8, Py-H6 and A-H2 protons appear in the aromatic region due to a pseudo aromatic effect created by lone pairs of electrons on the nitrogen atoms of the DNA bases. The purines guanine and adenine produce one and two singlets respectively, with the second adenine singlet due to the A-H2 proton. The pyrimidines are slightly more complicated with cytosine producing doublets in the aromatic region and the H1' region (5.5 ppm - 6.5 ppm). These are due to scalar coupling between C-H6 ('aromatic' proton) and C-H5. Thymine produces a singlet in the aromatic region, as well as a methyl peak appearing between 1-2 ppm.

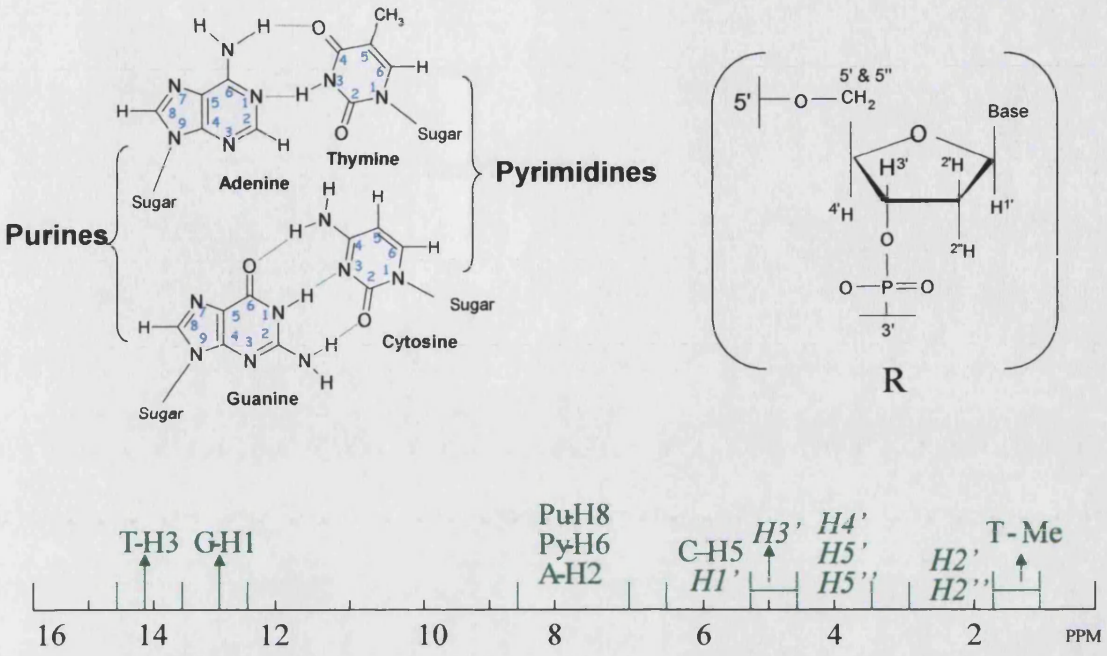


Figure 49: Numbering system for NMR and chemical shift values of deoxyribonucleotide base and sugar protons.

The one dimensional NMR spectrum of the 5'*d*(CTCATCAC).(GTGATGAG) duplex is shown in **Figure 50**. On obtaining the ^{1}H spectrum it is often possible to estimate the number of peaks in the aromatic region. For this duplex 24 peaks should be present (Gx4, Tx4, Ax8, Cx8), with the proton count integrating for 20 protons. In actual fact the peaks in this area are too overlaid to allow an accurate peak count, but the proton count integrates for 20 and is in a 1:1 ratio with the sugar H1' region, as expected. Four methyl peaks can also be seen between 1-2 ppm, indicating the presence of four thymine bases as expected. The spectrum is sharp, with a high signal to noise ratio, and no obvious impurities, and on the basis of this evidence it was possible to obtain 2D NOESY spectra in order to assign all protons.

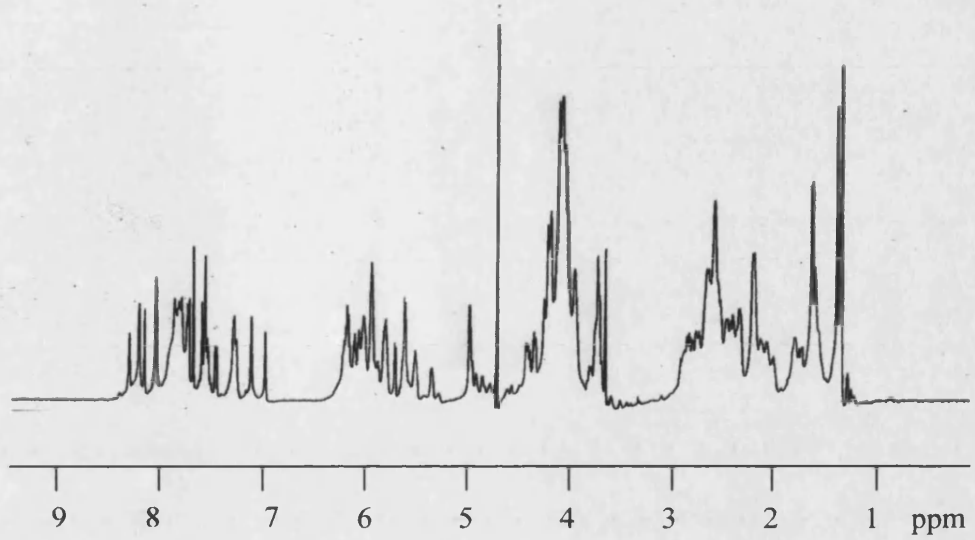


Figure 50: One dimensional Varian 400 MHz ^{1}H NMR spectrum of 5'*d*(CTCATCAC).(GTGATGAG)

2.6.2: The two-dimensional ^{1}H NOESY spectrum of the 5'd(CTCATCAC).5'd(GTGATGAG) duplex

In order to model a drug-DNA adduct it is necessary to use NOE methods to work out the distances between protons in the molecule. These distance restraints are then entered into molecular modeling programmes in order to work out the most likely orientation of the drug in the minor groove.

A diagram of the sequential connectivities between two DNA nucleotides can be found in **Figure 51**. Due to the orientation of bases in the DNA double helix any particular base can only 'see' into the 5' direction. That is to say that a base will display connectivities between its own protons, and also to the protons of its 5' neighbour, but will not 'see' back into its 3' neighbour. This means that a 5' terminal base will form connectivities only with its own protons, as it has no further 5' neighbour. This is important in the sequential assignment of DNA proton resonances, as it simplifies the spectrum immensely.

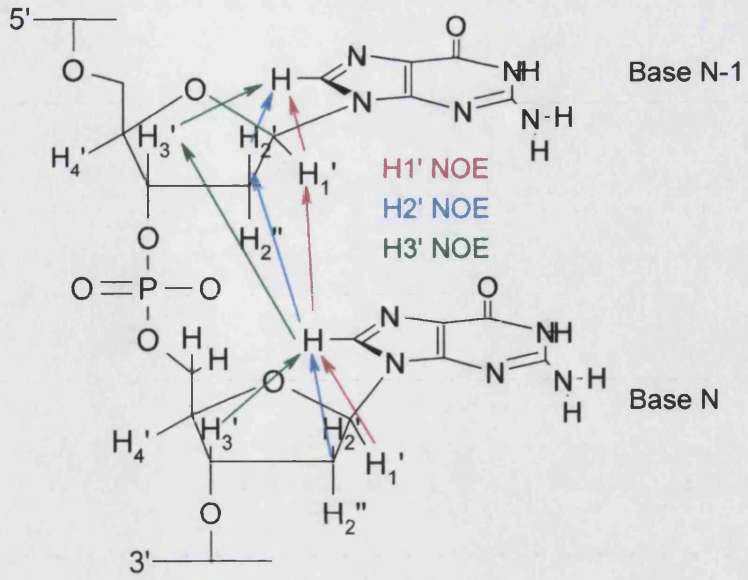


Figure 51: Diagram showing connectivities for two consecutive DNA bases

The entire 2D NOESY spectrum for the 5'*d*(CTCATCAC).(GTGATGAG) duplex is shown in **Figure 52**. It produced a set of well-resolved NOE off-diagonal peaks, with adequate signal to noise ratio for sequential assignment to begin. Each dimension of the spectrum was referenced to the H_2O peak at 4.71 ppm. In order for observable NOE peaks to be present in the spectrum the maximum ^1H-^1H distance must be within 5 Å. This means that only NOEs between protons within a nucleotide (intra-nucleotide) or between neighbouring nucleotides in the same strand (inter-nucleotide) are close enough to result in observable NOEs. Therefore each oligonucleotide strand gives rise to a cross-connectivity network between ribose and base protons of the constituent nucleotides. If sequences are self-complementary the cross-connectivity networks are superimposed upon each other and the spectrum is greatly simplified due to its symmetry. In this case the sequence was non-complementary however, so the proton resonances are different for each strand. The intense diagonal line of resonances across the spectrum is a projection of the one-dimensional spectrum previously discussed. On each side of the diagonal is a symmetrical arrangement of cross peaks resulting from dipole-dipole induced cross-relaxation. An NOE between two protons close in space will result in a contoured peak, the x and y coordinates of which indicate the chemical

shifts of the protons. In order to assign the NOESY spectrum a combination of techniques were used, all of which have been previously documented.[189-192]

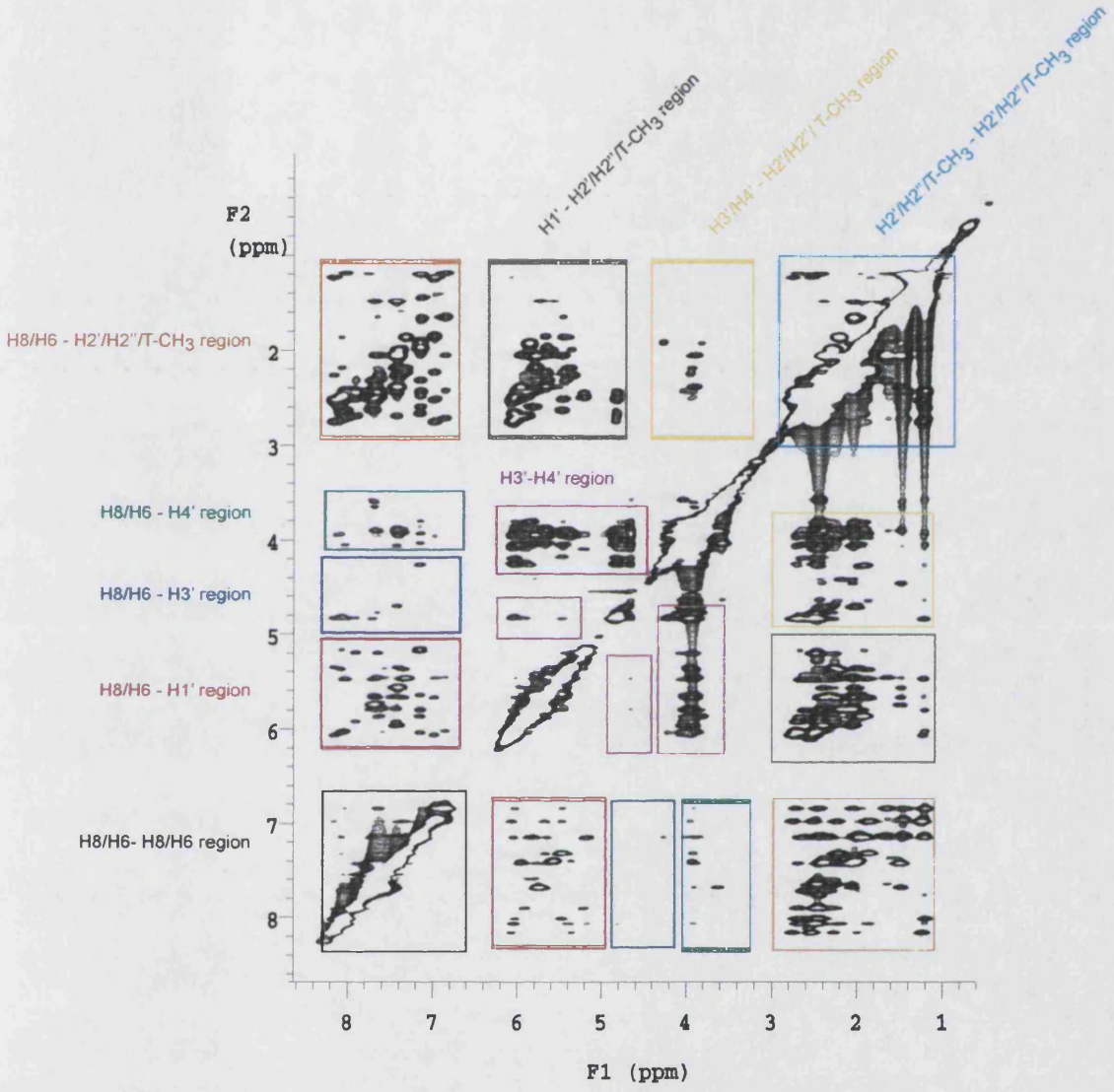


Figure 52: Full NOESY Spectrum for 5'*d*(CTCATCAC).(GTGATGAG)

2.6.3: Cytosine H5 protons (C-H5) (**Figure 53**)

Cytosine H5 protons are in such close proximity to the C-H6 protons that they produce extremely intense cross peaks in the NOESY spectrum, as well as COSY peaks in the COSY spectrum. This immediately provides a location for the C-H6 proton resonances and makes a good starting point in the assignment procedure. As shown in the 1D spectrum, C-H5 protons are found in the region of 5-6 ppm. When the H6/H8 – H1' region of the 2D NOESY spectrum (**Figure 53**) is studied it is possible to identify four extremely intense cross peaks at 5.20 ppm, 5.46 ppm, 5.56 ppm and 5.76 ppm in the F2 dimension that correspond to the C-H5 protons of the four cytosines in the oligonucleotide sequence. In addition to the COSY resonance produced by the coupling of C-H5 to C-H6 each C-H5 peak correlates directly into its own C-H1' and C-H6 protons in the F2 dimension. At 3.9 Å the C-H5 is also close enough to the aromatic base proton on the 5' side to produce an NOE connectivity in the F1 direction, thereby identifying the chemical shift of the aromatic base proton of the 5' flanking neighbour. At this stage it is not possible to identify which resonance belongs to which cytosine.

2.6.4: Thymine CH_3 protons (**Figure 54**)

The methyl protons of thymine produce a set of characteristic intense cross-peaks in the far up-field region of 1-2 ppm. If this region is examined in the 2D NOESY spectrum four intense pairs of peaks can be seen along the F2 dimension at 1.18 ppm, 1.20 ppm, 1.22 ppm, and 1.46 ppm. The most intense of each pair of peaks corresponds to the strong interaction (distance 2.9Å) between T-CH_3 and T-H6 of the same base. Therefore the T-H6 shifts of the four thymines in the sequence can be assigned as 6.86 ppm, 7.19 ppm, 7.00 ppm and 7.48 ppm respectively. The less intense peaks are due to an NOE between the T-CH_3 and the 5' flanking aromatic base proton. It is therefore possible to assign the four flanking bases' H8/H6 chemical shifts as 8.04 ppm, 7.77 ppm, 8.18ppm and 7.70 ppm. The two most down-field shifts correspond to adenine bases, which appear the most down-field due to increased ring-current effects. The further up-field peaks are due to guanine or cytosine bases. On examination of the DNA sequence it can be seen that there are indeed two incidences of thymines flanked by adenine bases (T5 and T13) while T2 and T10 are flanked by cytosine and guanine respectively. Peaks

from T-CH_3 to flanking aromatic protons are less intense due to the greater distance between these protons (3.8 Å). In this case, as with the cytosine H5 protons, it is not possible to further distinguish which thymine is which from this information.

2.6.5: The H8/H6 to H1' deoxyribose proton region. (**Figure 53**)

This region overlaps with the cytosine H5 region, spanning around 1.5 ppm from 5.1-6.3 ppm. Resonances in this region are due to the inter- and intra-nucleotide interaction between each aromatic H6 or H8 protons with the H1' proton of its own sugar, and also with that of its 5' flanking neighbour. The H8/H6 chemical shifts of the four cytosines and the four thymines are already known from examination of the cytosine H5 and thymine CH_3 regions. It is now possible to examine the connectivities of these protons with their neighbours in order to determine which resonance belongs to which base.

To begin with the thymine resonances at 7.00 ppm it can be seen that at this chemical shift there are a pair of peaks (5.66 ppm and 6.05 ppm) that line up in the vertical direction. These are due to the thymine H6 interaction with its own H1' proton and that of its 5' flanking neighbour. Looking horizontally along the F1 dimension from each of these resonances they are found to line up with two peaks at 7.35 ppm and 8.18 ppm. From examination of the cytosine H5 region and the thymine CH_3 region it is already known that these peaks correspond to a cytosine and an adenine respectively. The thymine in question then, is flanked by a cytosine and an adenine and can only be T5. The cytosine and adenine resonances can similarly be attributed to C6 and A4. Hence the H6/H8 and H1' resonances for A4 and T5 are now known. The C6 resonance is found to line up with another peak in the F2 dimension at 5.37 ppm, and this can therefore be assigned to the C6-H1' resonance. A chain of connectivity now exists from C6-T5-A4. It is now possible to continue to "walk" from resonance-resonance in this manner in both directions thereby identifying the other protons in the 5'*d*(CTCATCAC) strand. When C1 is reached it is important to note that this resonance has no other connectivity in the F2 dimension. This is because the C1 residue is at the 5' terminal and therefore can see only its own H1' sugar proton, as it has no 5' flanking neighbour.

The 5'*d*(GTGATGAG) strand can be assigned in the same manner, but is slightly more difficult due to the presence of overlying signals, namely G9-H1' and C1-H5. The chemical shift of G9-H8 is known from identification of T10 in the thymine CH_3 region (discussed previously). It is evident from the size of peaks and the number of contours that some peaks are overlaid, and using the information discussed as well as conformation from other regions of the NOESY spectrum (to follow) it is possible to tentatively identify H6/H8 and H1' resonances for all bases (**See Table 4**).

	C1	T2	C3	A4	T5	C6	A7	C8
H6/H8	7.70	7.48	7.42	8.18	7.00	7.35	8.10	7.19
H1'	5.67	5.96	5.34	6.05	5.66	5.37	6.03	5.86

Table 4: Chemical shifts in ppm for aromatic and H1' protons for the duplex 5'*d*(CTCATCAC).(GTGATGAG), C strand

	G9	T10	G11	A12	T13	G14	A15	G16
H6/H8	7.77	7.19	7.73	8.04	6.86	7.66	7.92	7.44
H1'	5.80	5.64	5.48	6.01	5.46	5.19	5.92	5.78

Table 5: showing chemical shifts in ppm for aromatic and H1' protons for the duplex 5'*d*(CTCATCAC).(GTGATGAG), G strand

As discussed previously, this region is also home to adenine H2 resonances. The position of adenine H2 protons in the centre of the minor groove allows close contact with protons within its own strand, and also NOE inter-strand connectivities. Each A-H2 can 'see' into its own deoxyribose H1' proton, into the thymine H1' proton to which it is Watson-Crick base paired, and also into the H1' proton of the 3' neighbour of that thymine. In this duplex the A-H2 cross peaks are very weak, but they will be discussed further in section 2.7.1.

2.6.6: The H6/H8 to H2' and H2'' region. **(Figure 54)**

This region spans from 1.8 – 3.0 ppm. It is used to confirm shifts assigned in the H8/H6 to H1' region, and clear up discrepancies regarding any overlaid peaks. As H6/H8 shifts have been suggested for all bases at this point it is possible to find all peaks coupling to each 'aromatic' base proton at any given chemical shift. At each H6/H8 shift four peaks should be present in the H2' and H2'' region, two of these correspond to the interaction between the base H6/H8 proton with its own H2' and H2'', while the other two are due to the interaction of the H6/H8 proton with the H2' and H2'' protons of its 5' flanking neighbour. As with the aromatic to H1' region, the number of peaks at the terminal base is halved due to the fact that this base can only interact with its own protons, as it has no 5' flanking neighbour.

To assign this region a pair of parallel lines are drawn from the H6/H8 resonances, in the F1 direction. These lines should intersect another pair of peaks at the chemical shift of the 5' flanking neighbour. These are due to the interaction of the H2' and H2'' of the first base with the H8/H6 proton of the neighbour. A clear example of this is shown between T5 and A4 in the region in question. From the assignment of the aromatic to H1' region it can be seen that the H8 resonance of the T5 base is at 7.12 ppm. At this resonance in the H2' and H2'' region four clear peaks can be seen. Parallel lines drawn along the F1 direction connect these peaks to four more peaks – one pair at the A4 chemical shift, and one pair corresponding to the shift value for C6. This confirms the H6/H8 proton assignments for these bases and also allows assignment of the T5-H2', H2'' and the A4-H2', H2'' peaks. These are assigned chemical shifts of 1.99 ppm, 2.38 ppm and 2.66 ppm, 2.90 ppm.

The H2' and H2'' can be distinguished between owing to the size of the NOE cross-peak with their own H8/H6 proton. The H2' proton is situated approximately 2.3Å from the 'aromatic' proton whereas the H2'' proton is further away at approximately 3.6 Å (distances based on UCSF-*Chimera* molecular models).[193] The H2'-H8/H6 cross-peak therefore is more intense than the H2''-H8/H6. In addition it is generally found that the H2' resonates down-field of the H2''. In this manner the H2 proton peaks are assigned for both strands (**See Tables 6** and **7**).

	C1	T2	C3	A4	T5	C6	A7	C8
H2'	2.38	2.40	2.26	2.74	2.24	2.18	2.66	2.64
H2''	2.04	2.04	1.98	2.47	1.84	1.86	2.46	1.90

Table 6: Chemical shifts (ppm) for the sugar H2' and H2'' protons in the 5'*d*(CTCATCAC).(GTGATGAG) duplex, C strand

	G9	T10	G11	A12	T13	G14	A15	G16
H2'	2.60	2.32	2.62	2.70	2.04	2.44	2.68	2.38
H2''	2.05	1.98	2.51	2.41	1.68	2.32	2.44	2.16

Table 7: Chemical shifts (ppm) for the sugar H2' and H2'' protons in the 5'*d*(CTCATCAC).(GTGATGAG) duplex, G strand

2.6.7: The H8/H6 to H3' region. **(Figure 55)**

This region is typically quite weak and full proton walks are rarely achieved. Due to water suppression in this region some cross peaks are also lost under the water suppression band. It is possible however, to confirm locations of some of the previously assigned peaks. For the duplex in question it was possible to confirm a fairly strong walk for the G strand, while several short chains of connectivity were discovered for the C strand. Missing chemical shifts were discovered using other regions of the spectrum and a tentative assignment was made for each proton. The data agreed with those assigned in previous regions and is shown in **Tables 8** and **9.**

	C1	T2	C3	A4	T5	C6	A7	C8
H3'	4.46	4.70	4.40	4.83	4.64	4.64	4.83	4.60

Table 8: Chemical shifts in ppm for H3' protons for the duplex 5'*d*(CTCATCAC).(GTGATGAG), C strand

	G9	T10	G11	A12	T13	G14	A15	G16
H3'	4.77	4.70	4.83	4.76	4.83	4.62	4.76	4.82

Table 9: Chemical shifts in ppm for H3' protons for the duplex 5'*d*(CTCATCAC).(GTGATGAG), G strand

2.6.8: The H8/H6 to H4' region

This region is home to connectivities between 'aromatic' and H4' protons. Each aromatic proton shows an NOE into its own H4' proton. This area is often heavily overlaid and unambiguous assignment of peaks is difficult.

2.6.9: The H8/H6 to H8/H6 region.

The H8/H6 – H8/H6 region is used to confirm locations of 'aromatic' protons and adenine H2 protons. Each Pu-H8 and Py-H6 proton correlates into the H8 or H6 proton of both its 3' and 5' neighbour. In the case of the 5'*d*(CTCATCAC).(GTGATGAG) duplex this region was too greatly ridged to be of much use. It will however, be discussed in relation to the 5'*d*$(CGATTAATCG)_2$ duplex.

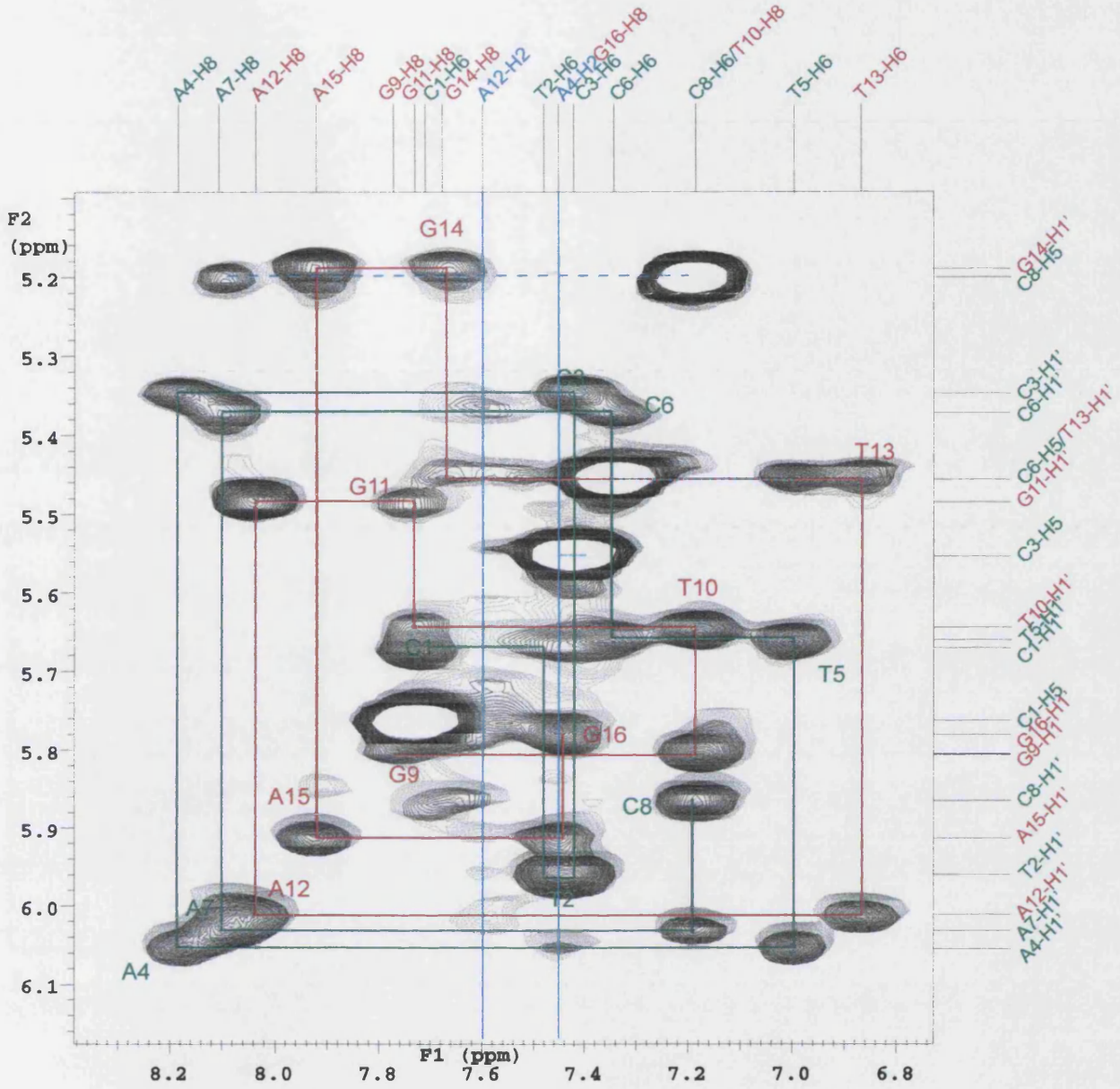


Figure 53: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CTCATCAC).(GTGATGAG), mixing time 200ms. C-H5 and H8/H6 – H1' region

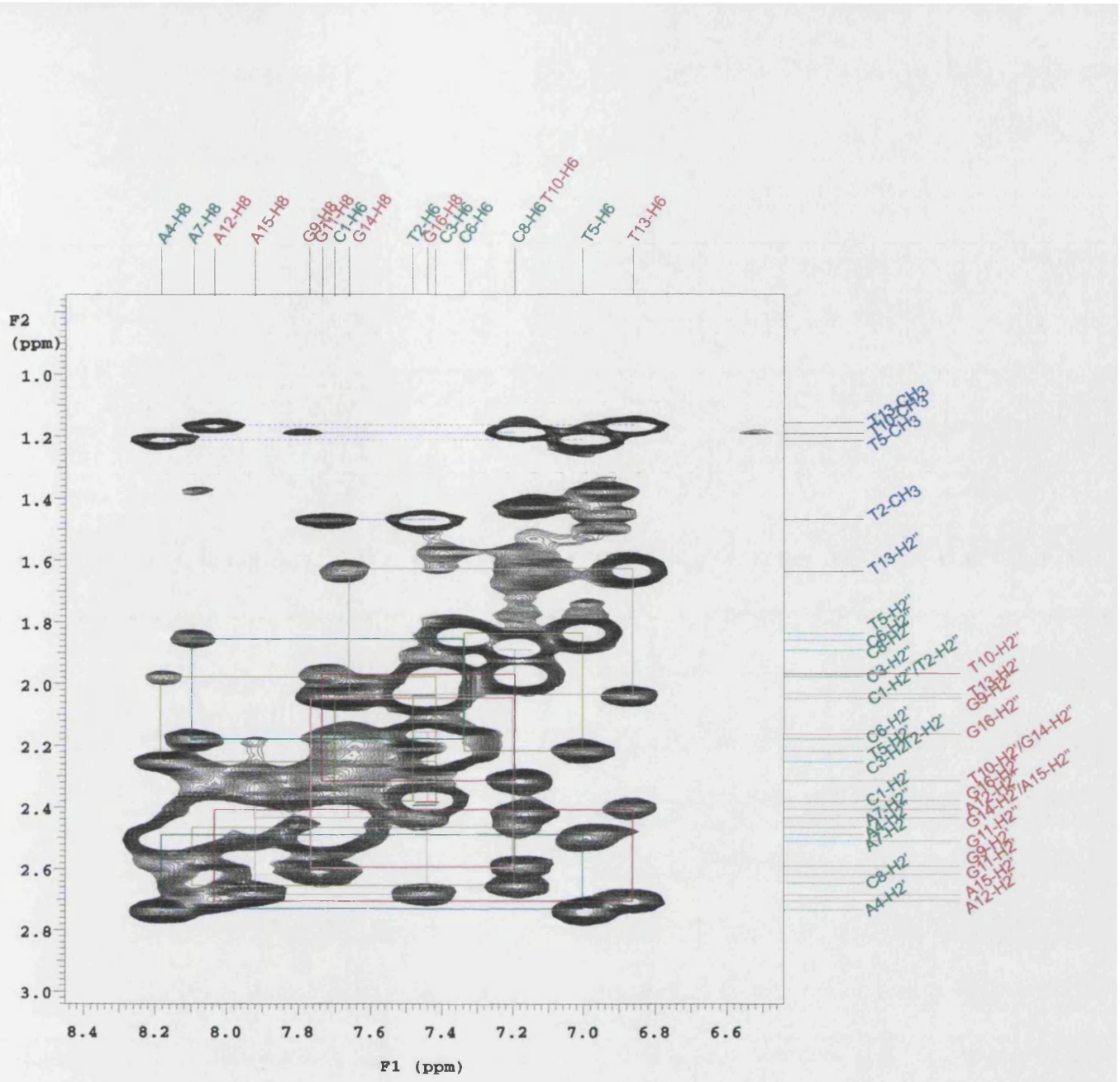


Figure 54: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CTCATCAC).(GTGATGAG), mixing time 200ms. Thymine methyls (blue) and H8/H6 to sugar H2' and H2'' walk, C strand = green, G strand=red

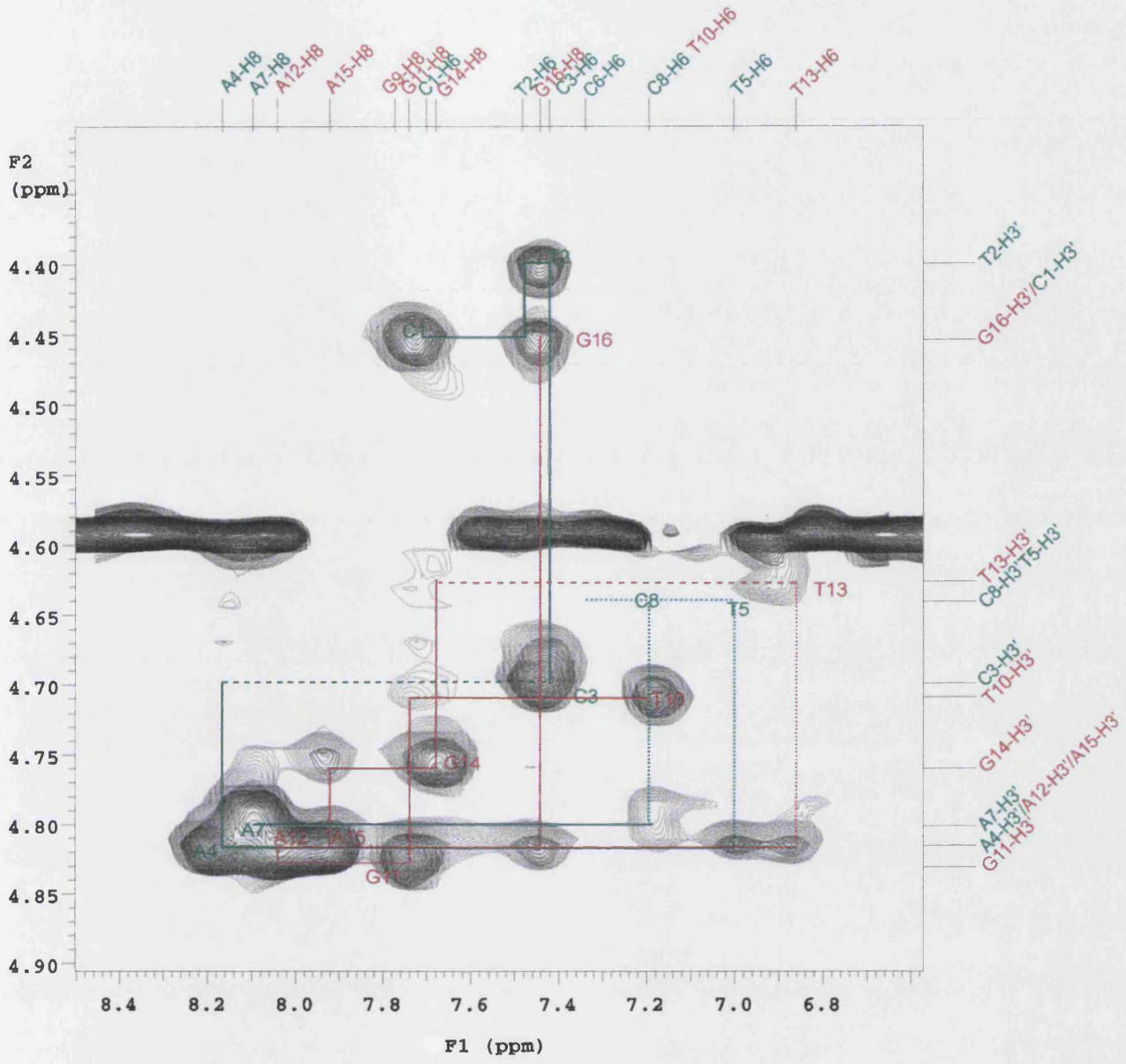


Figure 55: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CTCATCAC).(GTGATGAG), mixing time 200ms. The H8/H6 to sugar H3' walk. C strand = green, G strand = red

2.7: 2D NMR spectrum of the 5'*d*(CGATTAATCG)$_2$ duplex

The 2D NOESY spectra were assigned in a similar manner to that of the previous duplex, and so will not be discussed in depth here. It is evident that, as discussed, the spectrum is a lot less complicated and the assignments more clear in most regions. There are however, one or two areas that would benefit from further explanation.

2.7.1: Adenine-H2 assignment.

As previously discussed, A-H2 proton resonances are found at around 6.50 – 8.50 ppm in the aromatic region. They display connectivities to H1' resonances and therefore cross-peaks can be identified in the H6/H8 to H1' region of the spectrum (**Figure 57**). In the case of the palindromic duplex discussed here the simplification of the spectrum allows relatively easy assignment of these protons. Obvious peaks can be seen at 7.65 ppm that have not been included in the aromatic to H1' walk. If this chemical shift is studied in the H8/H6 to H2' and H2'' region (**Figure 58**) it can be observed that there are no cross-peaks in this region. The resonances seen at 7.65 ppm can therefore be attributed to the interaction of an A-H2 proton with sugar H1' protons. Other A-H2 resonances can be seen in the H6/H8 to H1' region underneath the A6-H8/A3-H8 chemical shift at 8.11 ppm. This is supported by the 1D spectrum (**Figure 56**), in which the peak seen at 8.11 ppm integrates for not two but three protons. Observation of the 1D spectrum also provides the clue for the location of the third A-H2 proton resonance, where an extra sharp singlet can be seen at 7.41 ppm - very close to the C1 resonance. This indicates that the third A-H2 resonance is probably situated underneath the ridged area around the C1 region of the 2D spectrum. To assign which A-H2 is which it is necessary to study their connectivities with other protons. As discussed earlier each A-H2 can 'see' into its own H1' proton, into its Watson-Crick paired thymine, and into the 3' neighbour of that thymine. In some cases a connectivity to the H1' proton of the 3' neighbour of the adenine H2 can also be observed. As the NOE peaks at 7.65 ppm in the F1 direction are the most clear these were studied first. These peaks display connectivities into A3-H1', T4-H1', T8-H1' and C9-H1' and the resonances are therefore due to A3-H2. The other peaks are heavily overlaid so are more difficult to assign. However, a clear resonance can be seen at 7.41 ppm in the F1 direction, which

shows clearly into the A7-H1' resonance at 5.96 ppm. This peak therefore can only be due to A7-H2, and the remaining resonances at 8.11 ppm can be assigned to A6-H2 by default, supported by the fact that underlying peaks can be observed at 5.52ppm, showing a connectivity into T5-H1'.

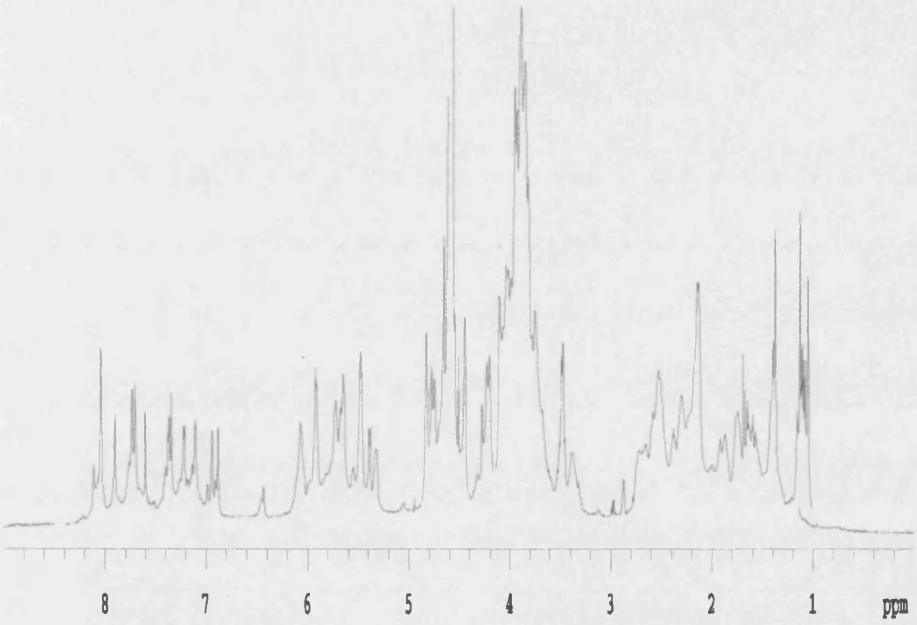


Figure 56: ^{1}H 600 MHz spectrum of 5'*d*(CGATTAATCG)$_2$

2.7.2: The H8/H6 to H8/H6 region

In this case the aromatic to aromatic region was much clearer than in the previous duplex, and it was possible to find correlations to back up earlier assignments in this region (**See Figure 59**).

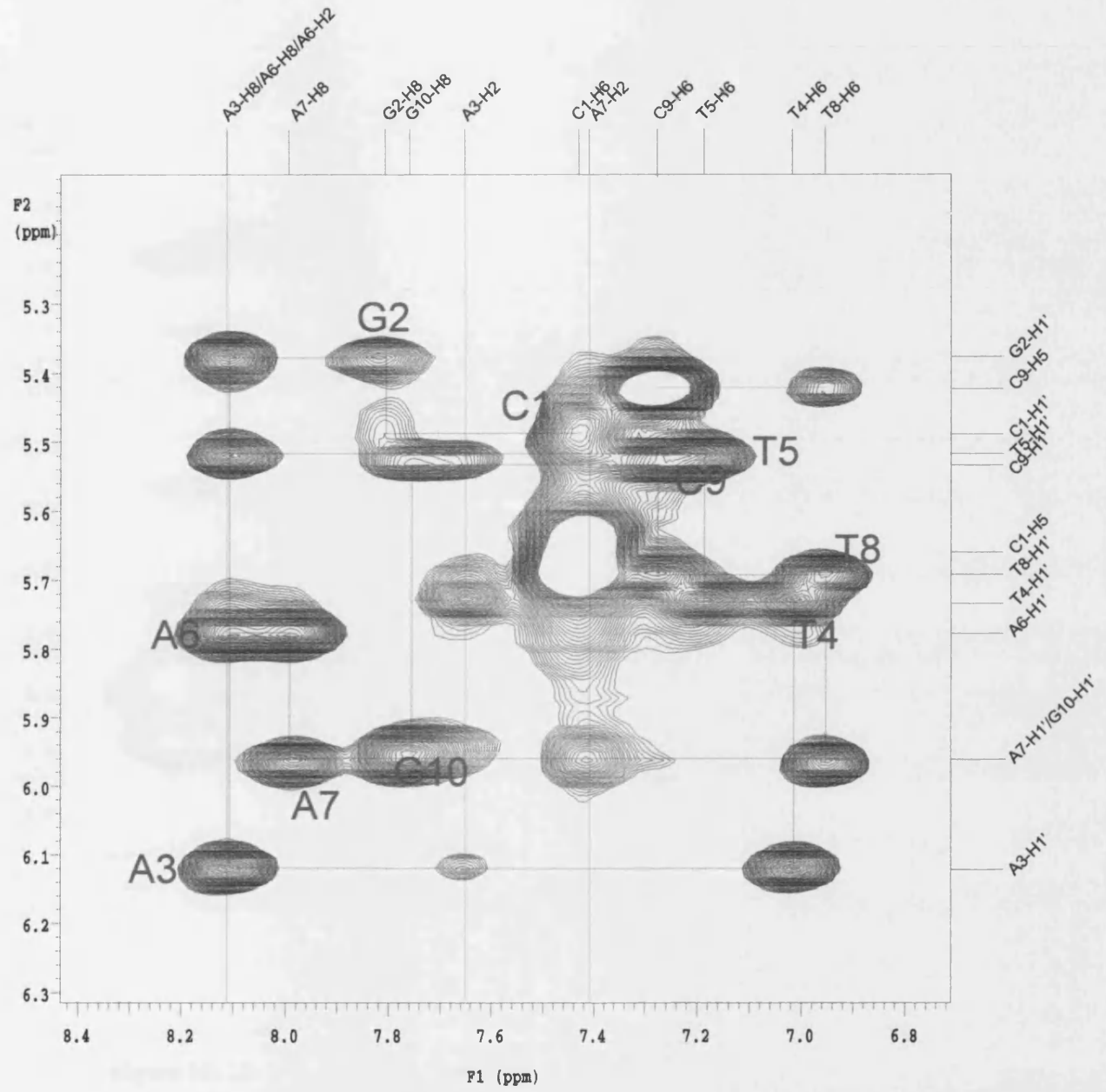


Figure 57: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CGATTAATCG)$_2$ mixing time 200ms H6/H8 to H1' region.

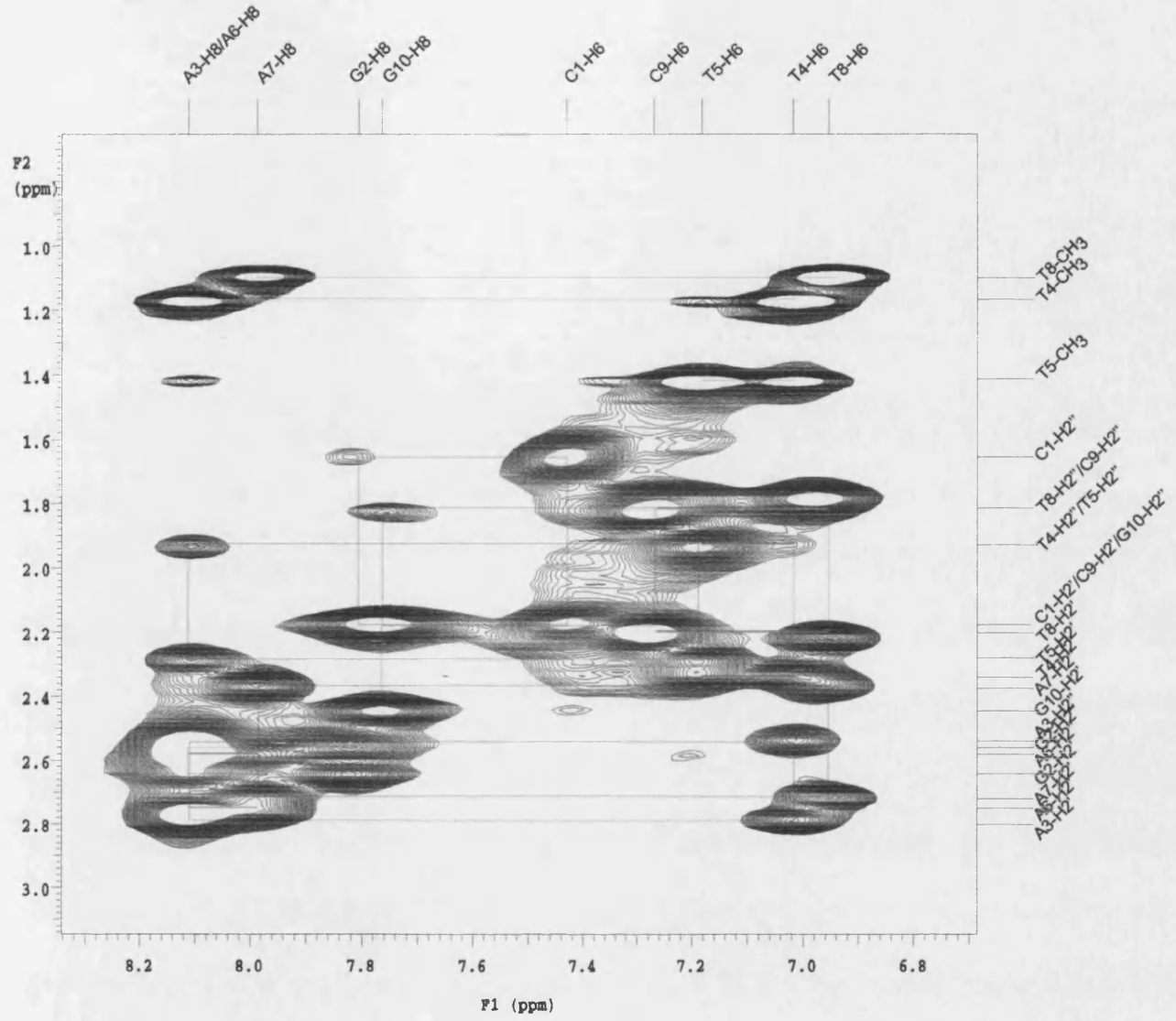


Figure 58: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CGATTAATCG)$_2$ mixing time 200ms H6/H8 to H2' and H2'' region

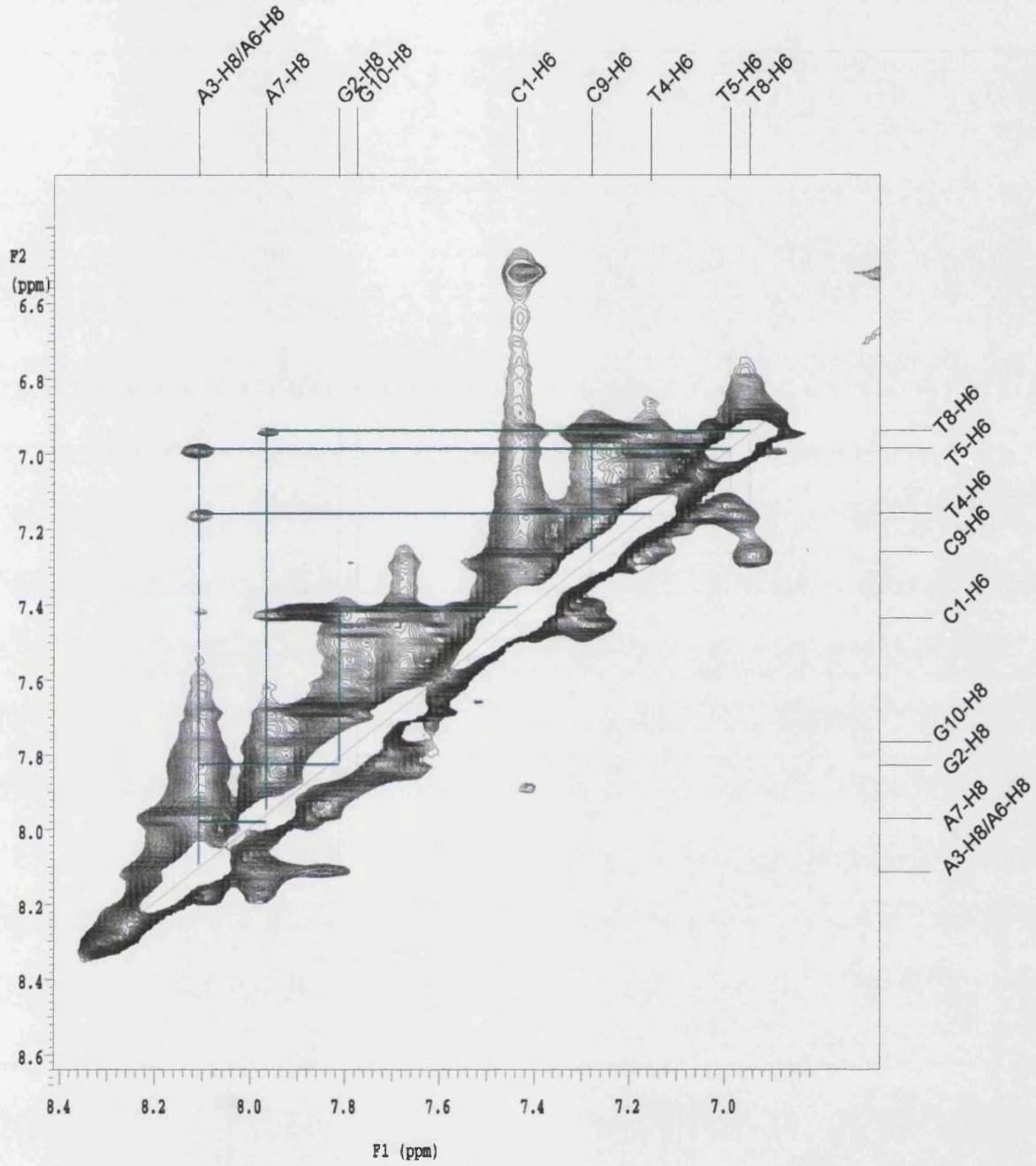


Figure 59: 2D NOESY Varian 600 MHz spectrum of 5'*d*(CGATTAATCG)$_2$ mixing time 200ms. H6/H8 to H6/H8 region

	H8/H6	H2/H5/CH_3	H1'	H2'1	H2'2	H3'	H4'
C1	7.70	5.76	5.67	2.38	2.04	4.46	3.90
T2	7.48	1.50	5.96	2.40	2.04	4.70	4.05
C3	7.42	5.56	5.34	2.26	1.98	4.40	3.96
A4	8.18	7.45	6.05	2.74	2.47	4.83	3.96
T5	7.00	1.24	5.66	2.24	1.84	4.64	3.96
C6	7.35	5.46	5.37	2.18	1.86	4.64	3.90
A7	8.10	7.67	6.03	2.66	2.46	4.83	3.96
C8	7.19	5.20	5.86	2.64	1.90	4.60	4.27
G9	7.77	-	5.80	2.60	2.05	4.77	3.90
T10	7.19	1.20	5.64	2.32	1.98	4.70	4.03
G11	7.73	-	5.48	2.62	2.51	4.83	4.16
A12	8.04	7.60	6.01	2.70	2.41	4.76	4.26
T13	6.86	1.19	5.46	2.04	1.68	4.83	3.89
G14	7.66	-	5.19	2.44	2.32	4.62	4.10
A15	7.92	7.60	5.92	2.68	2.44	4.76	4.23
G16	7.44	-	5.78	2.38	2.16	4.82	*

Table 10: Completed proton assignments for 5'*d*(CTCATCAC).(GTGATGAG). *=peak not found

	H8/H6	H2/H5/CH_3	H1'	H2'	H2''	H3'	H4'
C1	7.40	5.67	5.47	2.16	1.65	4.51	3.52
G2	7.80	-	5.36	2.66	2.55	4.83	4.17
A3	8.10	7.40	6.13	2.78	2.54	4.89	4.00
T4	7.00	1.17	5.71	2.32	1.79	4.71*	4.00
T5	7.16	1.43	5.52	2.28	1.93	4.71*	3.96
A6	8.10	8.11	5.76	2.74	2.58	4.33	4.28
A7	7.97	7.65	5.96	2.71	2.36	4.82	4.00
T8	6.94	1.09	5.70	2.22	1.78	4.90	3.96
C9	7.28	5.43	5.53	2.18	1.83	4.71*	3.90
G10	7.75	-	5.94	2.44	2.18	4.51	3.90

Table 11: Completed proton assignments for 5'*d*(CGATTAATCG)$_2$. *= peak not found hence assumed under water suppression band at 4.71ppm

2.8: Assignment of Drug-DNA Adduct spectra

After assignment of the duplex spectrum, a DNA duplex can be allowed to react with the DNA-binding drug. The resulting adduct is then subjected to further 1D and 2D NMR analysis. For the purpose of drug-DNA adduct assignment explanation the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct will be considered. Being a self-complementary sequence of only 10 base pairs this adduct provides a relatively simple spectrum for the purposes of an assignment explanation. **Figure 60** shows the numbering system of the duplex and the drug for the purposes of NMR assignment.

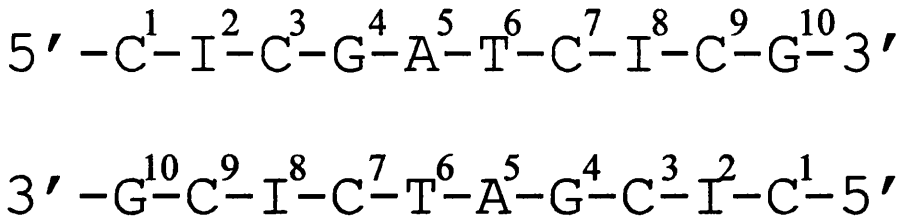


Figure 60: Numbering system for the 5'*d*(CICGATCICG)$_2$ duplex

2.8.1: The ^{1}H NMR spectrum

The ^{1}H NMR spectra of both the duplex and the adduct can be seen in **Figure 61**. These spectra are used to confirm that a reaction between the drug and the DNA has taken place. When the spectra are compared it is evident that different resonances are present in the adduct spectrum. As expected, the same regions of the spectrum are densely populated, but in particular new peaks can be observed at 1.80 ppm, 5.20 ppm and 6.60 ppm whilst peaks present in the duplex spectrum at 4.90 ppm, 7.50 ppm and 8.10 ppm are no longer present. This supports a drug-DNA reaction and suggests that no duplex DNA remains in the reaction mixture. Sharp singlets between 3 and 4 ppm are possibly due to the drug methyl group (SJG-H14). The adduct spectrum is also clean and well-resolved and so is deemed of a quality to allow 2D NMR experiments to take place.

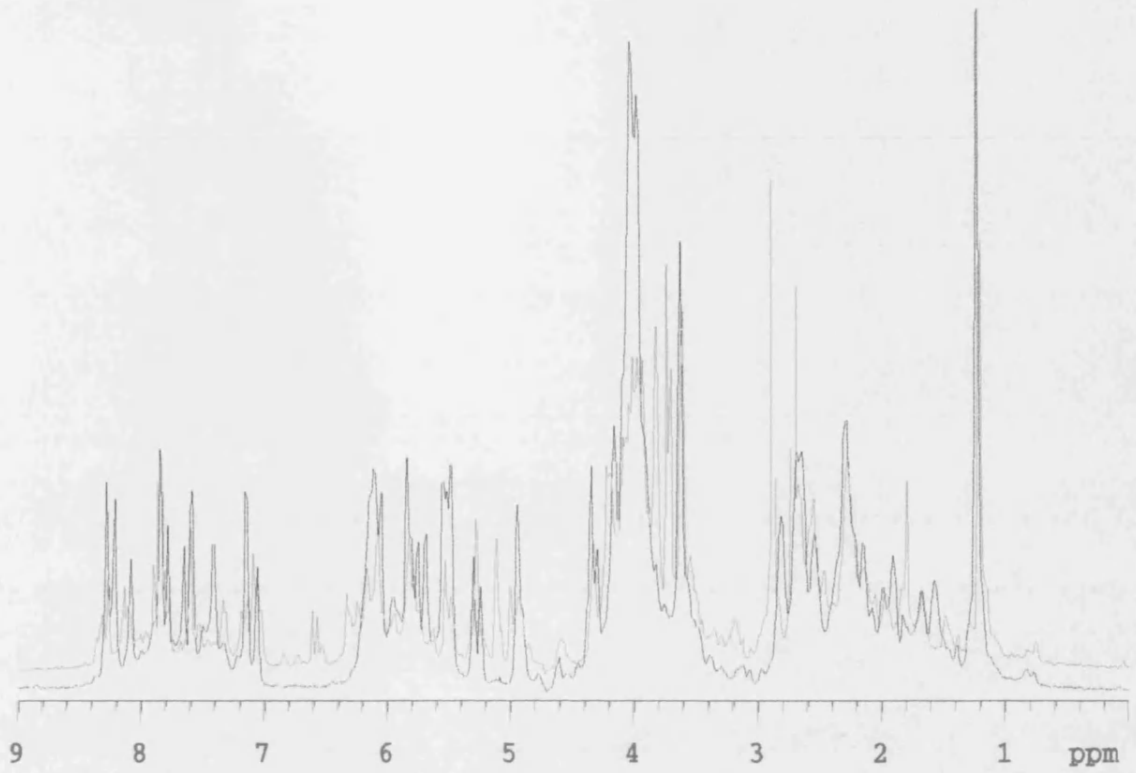


Figure 61: 600 MHz Varian ^{1}H NMR spectra of the 5'*d*(CICGATCICG)$_2$ duplex and the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct. The duplex spectrum is shown in black while the adduct is superimposed in red

*2.8.2: The 2D NOESY and COSY spectra of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct*

The NOESY and COSY spectra for the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct showing all drug resonances can be found in **Figure 65.** On comparison of the 2D NOESY spectrum of this adduct with that of the 5'*d*(CICGATCICG)$_2$ duplex it is confirmed that a drug-DNA reaction has taken place and that there is no residual duplex remaining (**See Figure 62**).

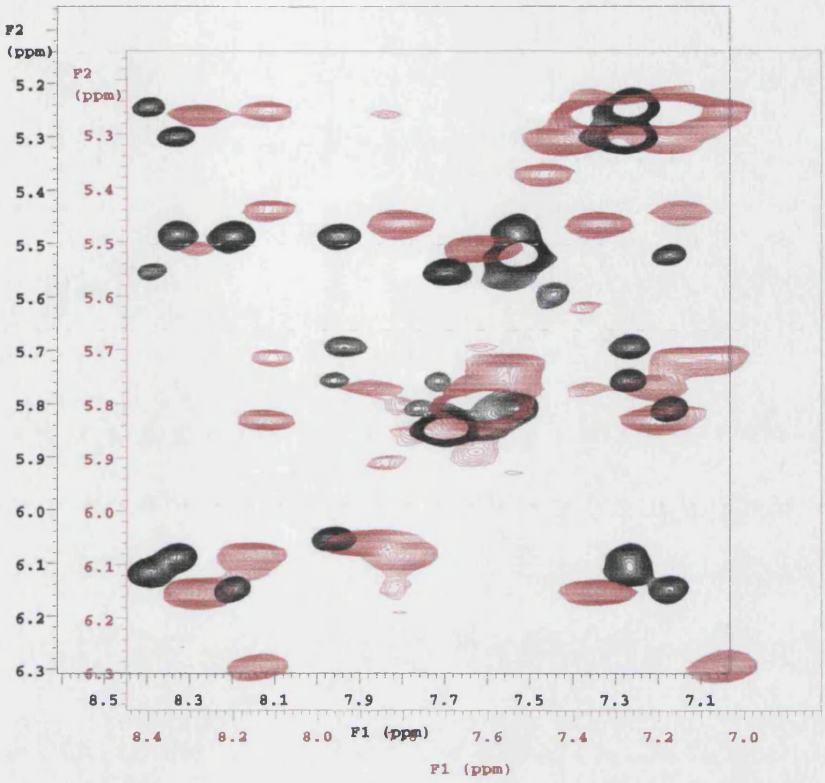


Figure 62: 600 MHz NOESY spectra of the 5'*d*(CICGATCICG)$_2$ duplex and the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, H8/H6-H1' region. The duplex spectrum is shown in black while the adduct spectrum is superimposed in red.

The sequential assignment procedure for the drug-DNA adduct begins in the same fashion as for the duplex, with assignment of the duplex protons based on the same connectivities as discussed in sections 2.5 and 2.6. It is usual to begin with the H6/H8 to H1' region and successful assignment of a complete 'walk' in this region confirms that the DNA has retained B-form character. The 2D NOESY expansion of this region fully assigned can be found in **Figure 63.**

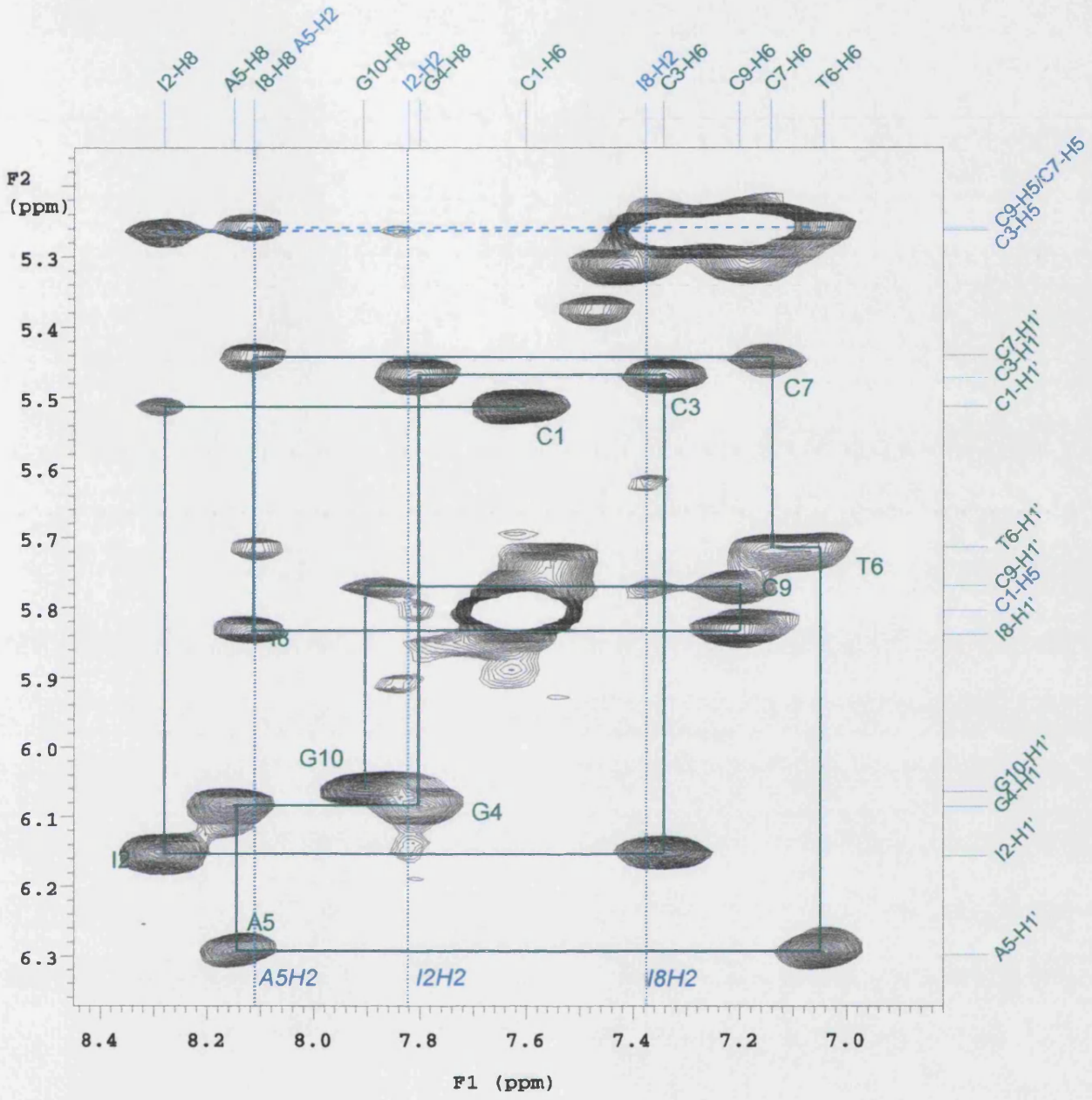


Figure 63: 2D NOESY spectrum of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, H8/H6-H1' region 600 MHz, 300ms mixing time

The presence of only one walk in this region confirms that the duplex has also maintained self-complementarity, and that only one adduct is present in the reaction mixture.

2.8.3: Assignment of drug resonances

Drug protons are found using intra-drug connectivities, which provide a 'walk' across the entire spectrum. The most useful connectivities present within SJG-136 are shown in **Figure 64.**

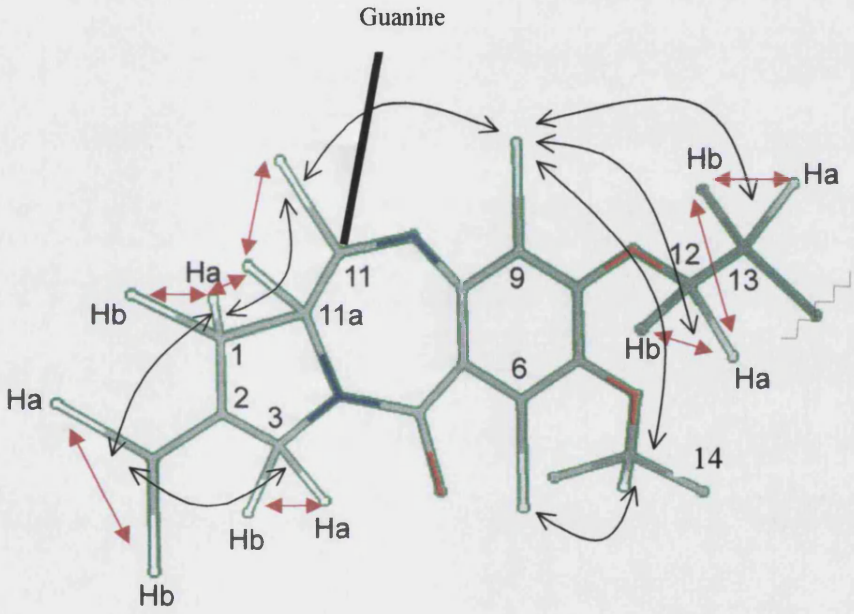


Figure 64: Pictorial representation of one end of the symmetrical covalent minor groove binder, SJG-136. The numbering system is shown as well as useful connectivities within the drug. COSY (scalar) couplings are shown in red with NOESY (dipolar) couplings represented in black

The assignment procedure for the drug protons of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct began with the SJG-H11 and SJG-H11a protons. These protons are $^3J_{HH}$ scalar coupled and so exhibit a strong COSY cross-peak. When the 2D COSY spectrum is examined a COSY peak can be observed in the H1' – H4' region (**Peak A, Figure 65**).

As H1' – H4' connectivities do not result in COSY interactions this peak can only be attributed to a drug interaction. Previous work on the similar 5'*d*(CICGATCICG)$_2$-DSB-120 adduct[181] located the SJG-H11 and SJG-H11a protons at a chemical shift of 5.75 ppm and 3.83 ppm respectively so it is likely that the COSY peak in question is the result of the SJG-H11-H11a interaction and as a result these protons are tentatively assigned chemical shifts of 5.54 ppm and 3.59 ppm respectively.

From the SJG-H11a resonance at 3.59 ppm it is possible to identify the SJG-H1a and SJG-H1b proton shifts by virtue of a COSY connectivity with SJG-H11a. In practise although both SJG-H1a and SJG-H1b are $^3J_{HH}$ scalar coupled to SJG-H11a only one COSY cross-peak is found. This is due to the dihedral angle between the two coupled protons. As discussed in section 2.4.4, if the angle is close to 90° the coupling constant becomes very small. In the case of the SJG-H11a-SJG-H1a/b connectivity the coupling between SJG-H11a – SJG-H1b would be expected to produce a COSY cross-peak, whereas the SJG-H11a – SJG-H1a coupling might not. If a line is drawn in the F1 direction from the H11a chemical shift of 3.59 ppm it bisects a strong COSY peak at 3.04 ppm (**Figure 65, Peak B**), thereby identifying SJG-H1a/b. The second SJG-H1 proton can be found by a COSY connectivity between SJG-H1a and SJG-H1b (**Figure 65, Peak C**) and is assigned a chemical shift of 2.51 ppm.

When vertical lines are traced in the F2 direction from the chemical shifts of SJG-H1a and SJG-H1b as expected they bisect a pair of NOESY cross-peaks at the SJG-H11 resonance of 5.54 ppm (**Figure 65, Peaks E & F**). At this point it is possible to tentatively distinguish between SJG-H1a and SJG-H1b on the basis of the strength of the interaction with SJG-H11. There is a slight difference in the distance of SJG-H1a to SJG-H11 (approximately 2.3 Å) and the longer distance between SJG-H1b and SJG-H11(approximately 2.9 Å). This results in a slightly stronger NOESY cross-peak between SJG-H1a and SJG-H11, and therefore SJG-H1a and SJG-H1b are assigned chemical shifts of 2.51 ppm and 3.04 ppm respectively. These assignments are supported by the presence of the COSY connectivity from SJG-H11a – SJG-H1a/b, which as expected, occurs as a result of the connectivity of SJG-H11a with SJG-H1b.

Continuance of the SJG-H1a and SJG-H1b lines in the F2 direction sees them bisect another pair of NOESY resonances at a chemical shift of 5.58 ppm (**Figure 65, Peaks**

G&H). This resonance can be assigned as SJG-H2a/b, and the lack of any other pairs of NOESY peaks in this region as well as a lack of any COSY connectivity that could be attributed to a correlation between the SJG-H2 peaks suggests that in this case SJG-H2a and SJG-H2b are overlaid.

Scanning of the SJG-H2a/b chemical shift in the F1 direction results in the observance of a pair of peaks at 4.00 ppm and 4.09 ppm (**Figure 65, Peak pair I**). These can be attributed to SJG-H3a and SJG-H3b although due to a lack of clear NOE resonances to surrounding DNA protons and the similarity in chemical shift of the two it is not possible to distinguish between the pair. This almost completes the assignment of the PBD subunit protons, leaving only SJG-H9 and the protons of the linker unit unassigned.

The SJG-H9 proton provides a lead into the linker protons by virtue of its connectivity with SJG-H12 and SJG-H14, and can be easily found due to its close proximity to the A5-H2 proton of the DNA duplex. When connectivities to the A5-H2 proton at 8.10 ppm are examined a large NOESY connectivity is observed at 6.55 ppm that is not due to a DNA connectivity (**Figure 65, Peak J**). This cross-peak can be attributed to the interaction of SJG-H9 and A5-H2 and hence the SJG-H9 proton is assigned a chemical shift of 6.55 ppm.

From the SJG-H9 resonance it is possible to identify SJG-H12a/b (both at the same chemical shift) by a NOESY cross-peak at 4.20 ppm (**Figure 65, Peak K**). SJG-H12a/b is $^{3}J_{HH}$ scalar coupled to SJG-H13a/b and so the connectivity results in a COSY interaction. The COSY cross-peak can be found at 2.30 ppm (**Figure 65, Peak L**), hence confirming the chemical shift of SJG-H13a/b. The lack of any other significant NOESY or COSY cross-peaks at the SJG-H12 chemical shift in this region leads to the conclusion that SJG-H13a and SJG-H13b are also found at the same chemical shift.

There now remain only two proton resonances to identify – SJG-H6 and SJG-H14. The connectivity between these would be expected to produce a large NOESY cross-peak in the DNA H8/H6 to H3' or H4' region. Such a cross-peak can be seen at coordinates 7.05 ppm/3.80 ppm and hence the SJG-H6 and SJG-H14 protons can be assigned shifts of 7.05 ppm and 3.80 ppm respectively.

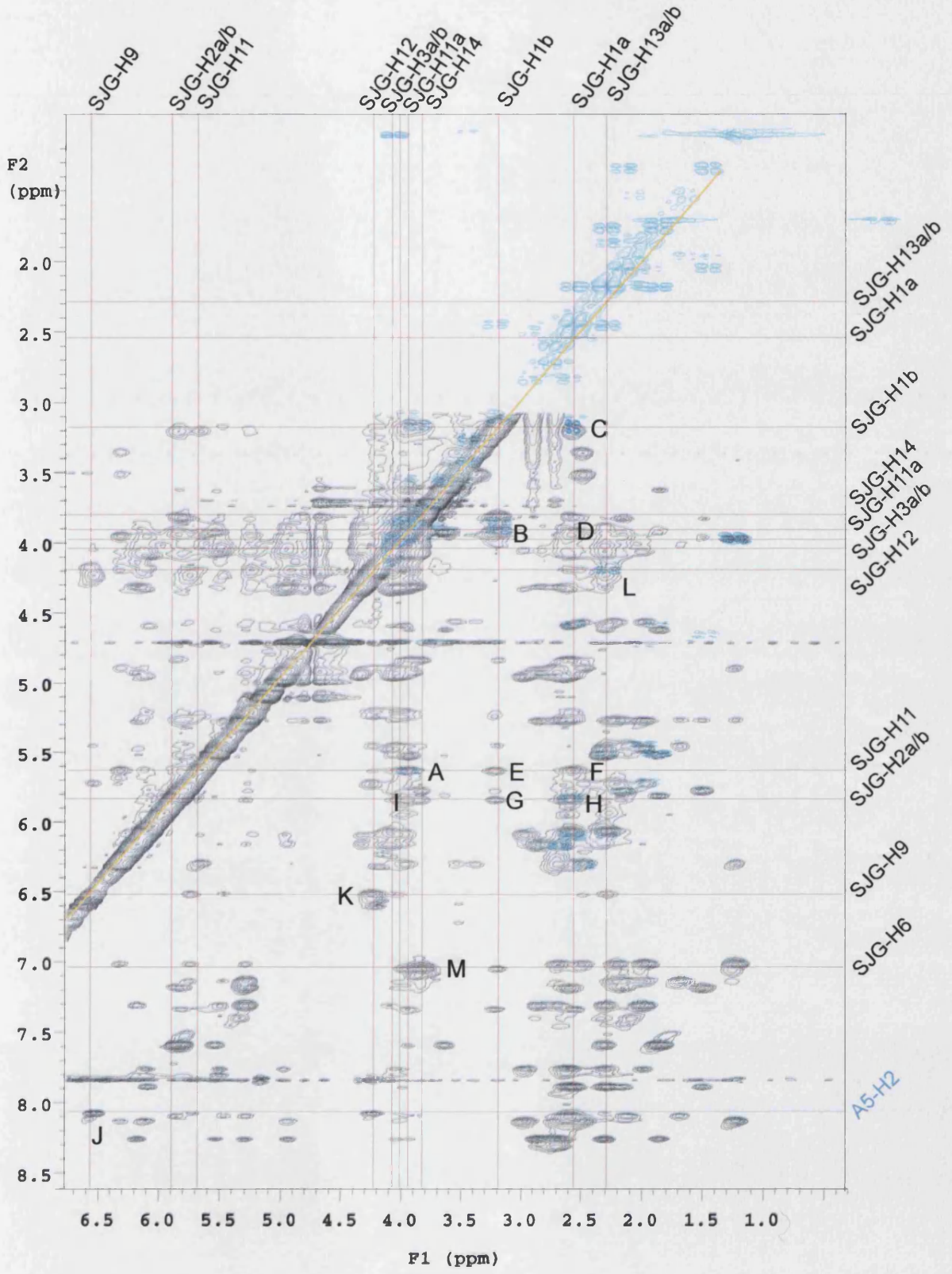


Figure 65: 600 MHz NOESY spectrum of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct 200 ms mixing time

On completion of assignment of drug protons in this way it is possible to confirm some assignments using NOE connectivities into the DNA duplex. Many of the drug-DNA connectivities are discussed in **Chapter 3**.

*2.8.4: Assignment of drug protons in the 5'*d*(CGATTAATCG)$_2$-Adozelesin Adduct*

As with the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, after ascertaining that the drug has reacted with the DNA and that there are no remaining duplex signals the first step in assigning the DNA-adozelesin adduct is to identify and assign all resonances belonging to the DNA. The drug proton walks are then identified in a similar manner using connectivities within the drug, as well as NOE connectivities to nearby DNA protons. The expected connectivities within an adozelesin molecule are shown in **Figure 66**.

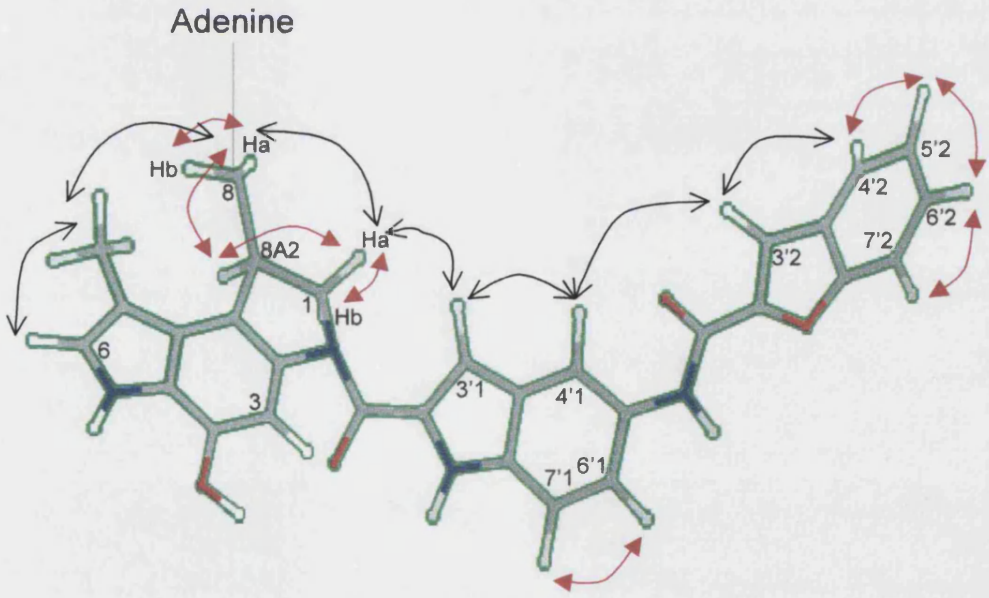


Figure 66: Pictorial representation of the covalent minor groove binder, adozelesin. The numbering system is shown as well as useful connectivities within the drug. COSY (scalar) couplings are shown in red with NOESY (dipolar) couplings represented in black

In the assignment of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct the drug methyl groups provided a good starting point due to the two large and fairly obvious NOESY cross-peaks in the aromatic-methyl region at 2.70 ppm and 2.85 ppm (**Figure 67, Peaks A & B**) corresponding to the interaction of the adozelesin methyl groups with the H6 proton. The Ado-H6 and Ado-methyl resonances were hence assigned shifts of 8.30 ppm and 2.85 ppm respectively (adduct 1), 8.30 ppm and 2.70 ppm (adduct 2). From the methyl proton in each case a cross-peak can be found in the F2 dimension corresponding to the interaction of this proton with the Ado-H8a/b protons (**Figure 67, Peaks C & D**). The Ado-H8a and Ado-H8b protons were found to be overlaid in both adducts, as only one cross-peak with the drug methyl group was observed for each adduct.

Ado-H8A2 was found in both adducts owing to a large NOE with Ado-H8a/b (**Figure 67, Peaks E & F**). There was no evidence of the expected COSY peak resulting from this interaction in either adduct, suggesting that in both cases the dihedral angle was close to 90°. Ado-H8A2 then provides a lead into the chemical shift of the Ado-H1a and Ado-H1b protons by way of a COSY connectivity with these (**Figure 67, Peaks G & H**). Once again, the presence of only one such COSY peak for each adduct confirms that, as with the Ado-H8a/b protons, the Ado-H1a and Ado-H1b protons are overlaid in both cases. The Ado-H1a/b protons then provide a connectivity with the Ado-H3'1 proton in the DNA H8/H6 – H4' region of the spectrum. The Ado-H3'1 proton chemical shift can be supported in the DNA H8/H6 - H8/H6 region of the spectrum by its connectivity with Ado-H4'1. However, care must be taken when confirming the location of Ado-H3'1 in this manner as there are many peaks present in this area.

The protons of the benzofuran tail and the H6'1-H7'1 vicinal coupling are found by virtue of strong COSY peaks in the DNA H8/H6 – H8/H6 region. There are no DNA protons that give rise to COSY peaks in this area, hence any COSY peaks detected can be safely attributed to drug protons. The Ado-H6'1 – Ado-H7'1 is an isolated spin system and so can usually be attributed to any peak that does not align with any others, while the remaining vicinal couplings around the benzofuran tail give rise to a connectivity network. In the case of the 5'*d*(CGATTAATCG)$_2$-adozelesin adducts assignment of this region was difficult owing to overcrowding. In addition many protons were situated at very similar chemical shifts – leading to the COSY cross-peak

being situated on the diagonal. The protons were assigned using a mixture of techniques including connectivities with DNA protons such as the adenine H2 protons and connectivities with thymine and adenine exchangeable amino H3 protons, seen only on acquisition of additional 2D spectra in a water solvent. The assigned region can be seen in **Figure 68**.

Differentiation between the two adducts was based on connectivities into the DNA backbone. From these it is possible to link the drug walk into a distinct DNA walk. Evidence for the Hoogsteen or Watson-Crick nature of each adduct will be subsequently discussed in **Chapter 5**.

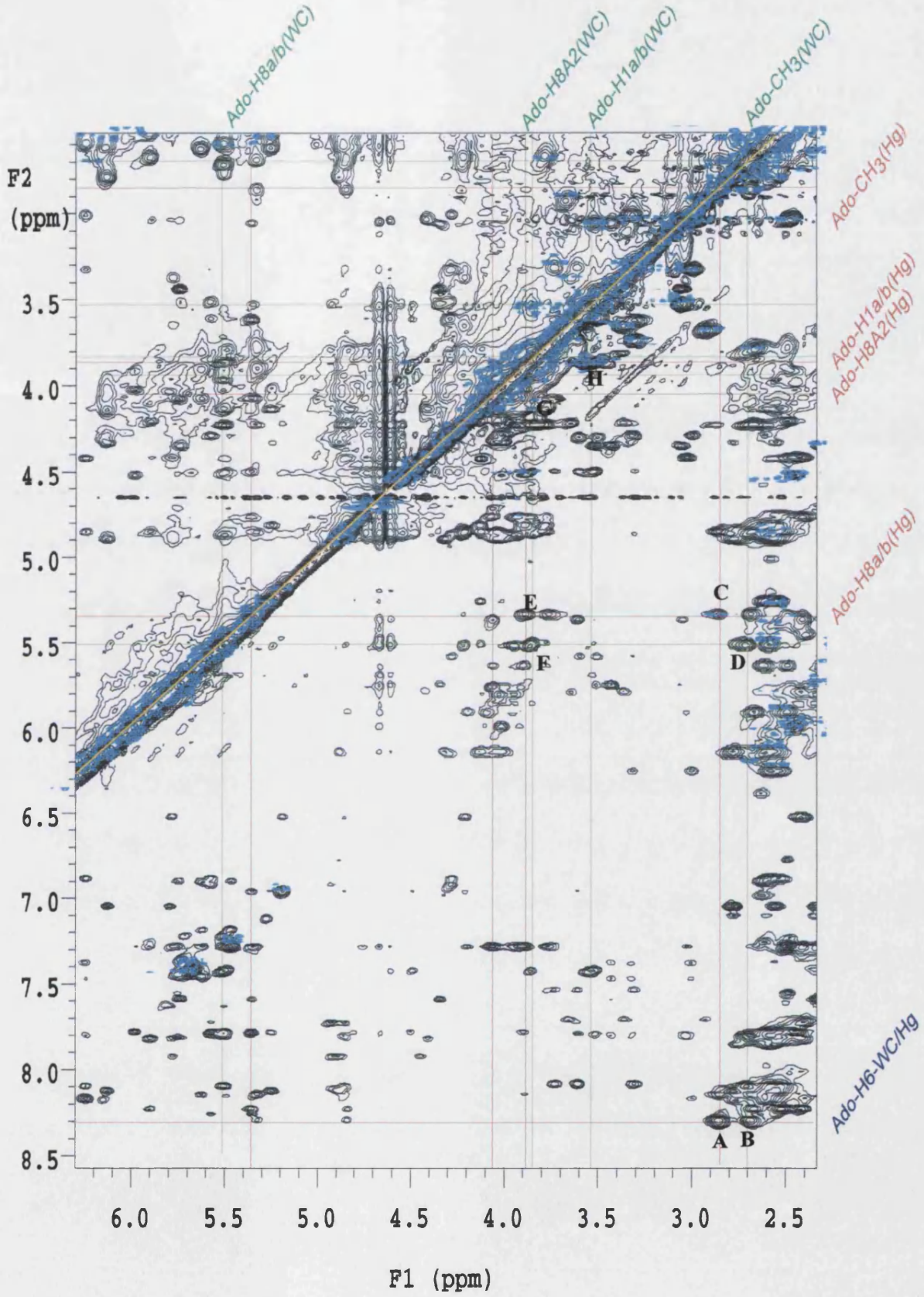


Figure 67: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. NOESY mixing time 200ms. Black=NOESY, blue = COSY. Green lines = Watson-Crick assignments, red lines = Hoogsteen assignments

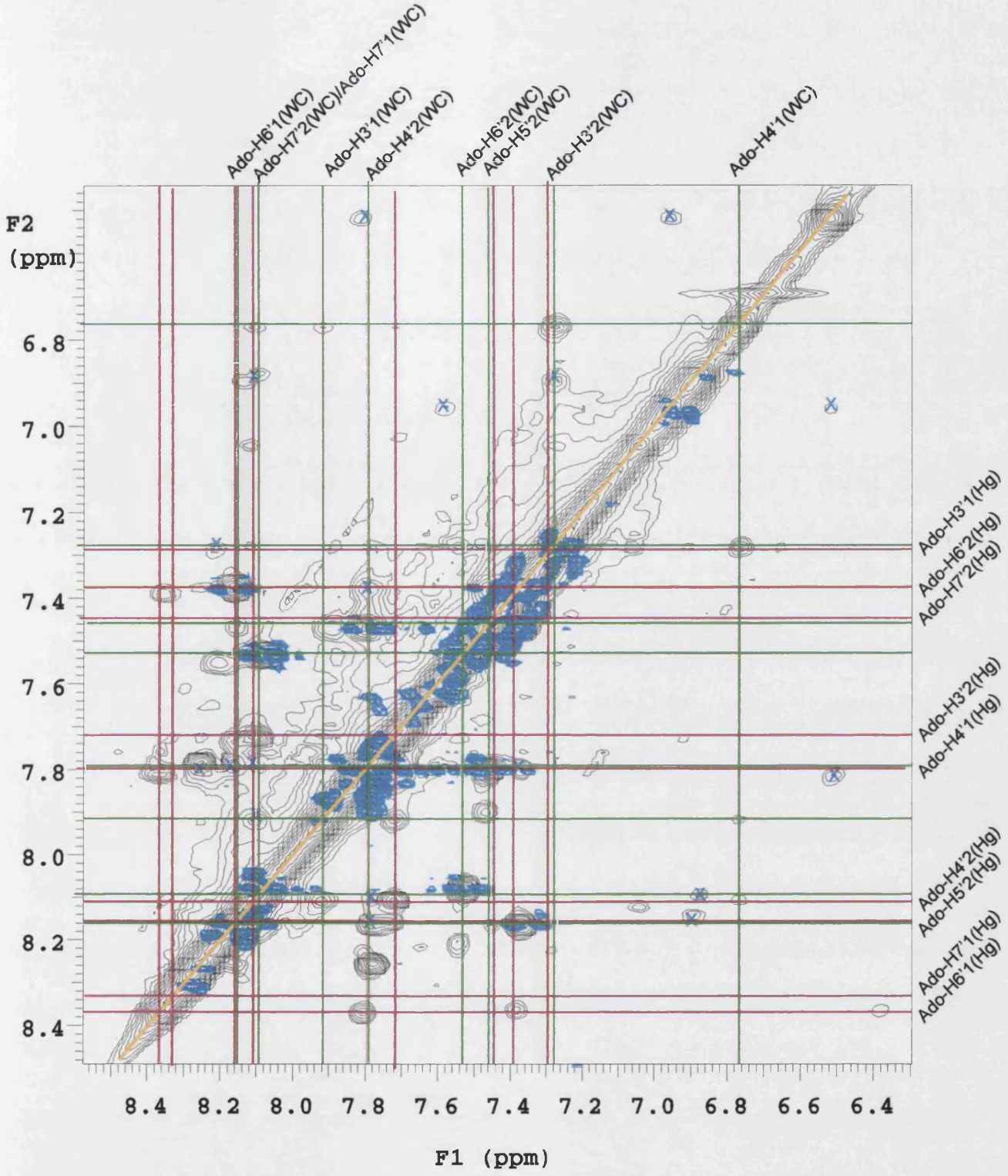


Figure 68: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. Black = NOESY, Blue= COSY. NOESY mixing time 200ms. Expansion of the H6/H8-H6/H8 region. Green lines = Watson-Crick assignments, Red lines = Hoogsteen assignments

CHAPTER 3: RESULTS AND DISCUSSION

THE 5'*d*(CICGATCICG)$_2$-SJG-136 INTER-STRAND ADDUCT

As discussed in **Chapter 1**, SJG-136 (**Figure 69**) is an *exo* unsaturated PBD dimer, linked through the C8 positions of the PBD subunits by a flexible propyl chain. It was designed to form inter-strand cross-links with B-form DNA. It was postulated that the drug would react similarly to DSB-120 (**Figure 69**) but with increased potency, as it was known that PBD monomers with unsaturation at the C2 position were more biologically effective.

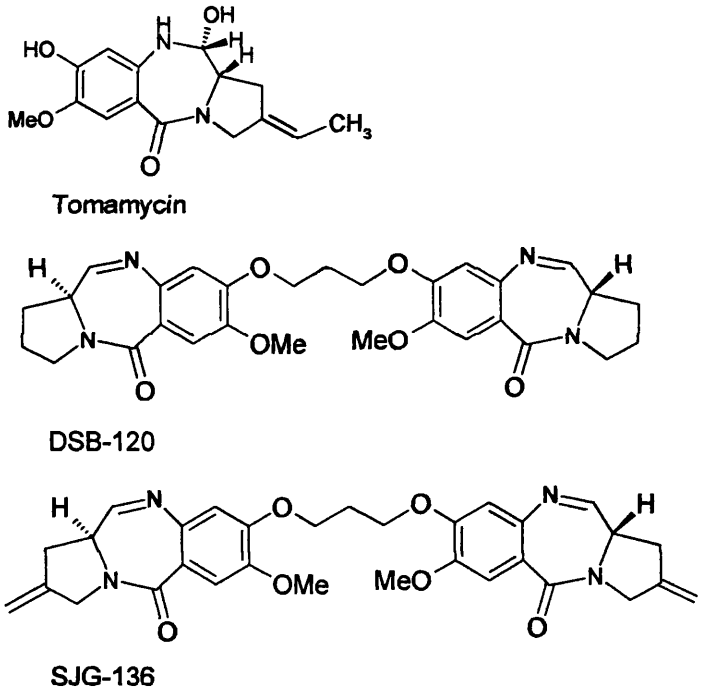


Figure 69: Structures of SJG-136, DSB-120 and tomamycin

The increased potency of the monomers is thought to be a result of lower electrophilicity at the N10-C11 position. On a concentration basis SJG-136 was found to show a 10-fold greater efficiency than DSB-120. Molecular modeling has shown an SJG-136 adduct to maintain the structure of B-form DNA and both SJG-136 and DSB-120 are easily accommodated in the minor groove. The lack of exposure beyond the

DNA periphery leads to an observed resistance to DNA repair enzymes, which function by tracing helical perturbation or distortion in DNA.[72] Calculated binding energies suggest that SJG-136 should form a more favourable adduct than DSB-120 as introduction of an *exo*cyclic methylene at C2 causes the C ring to stiffen and become more planar and hence a better fit in the minor groove.[73] In this study an SJG-136 adduct has been formed and compared with an identical duplex sequence which had formed adducts with DSB-120 and tomamycin in an earlier study.[181]

3.1: Design of Duplex

The 5'*d*(CICGATCICG)$_2$ duplex was synthesized using standard methods and was allowed to react for 16 hours with 5 mg of SJG-136 as detailed in **Chapter 2.** The duplex was identical to the duplex used in the study of DSB-120, in order to allow a direct comparison to be made. From the 1D ^{1}H NMR spectrum it was deduced that the reaction had proceeded to completion and 2D COSY and NOESY experiments were performed on the resulting adduct. Prior to performing molecular modeling experiments on the adduct it was necessary to determine the complexity of the cross-linked species formed, identify the sites of covalent linkage between SJG-136 and DNA and determine the stereochemistry and orientation of the drug within the DNA duplex.

3.2: Symmetry of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct

The one-dimensional ^{1}H-NMR non-exchangeable spectra for both the duplex and the adduct can be seen in **Figures 70 & 71.** The spectrum was clean with well-resolved signals. The resonances have been assigned by two-dimensional NMR spectra employing through-bond COSY and through-space NOESY connectivities. The chemical shift assignments for the nucleotide and drug protons in the duplex and the adduct are listed in **Tables 12 & 13**. There were no duplex resonances remaining after reaction with SJG-136 confirming that the reaction has proceeded to completion. The complete 2D NOESY and COSY spectra with drug resonances assigned can be seen in **Figure 72**, while an expansion of the H6/H8 – H1' region of the adduct NOESY and COSY spectra can be seen in **Figure 73**. Four strong COSY peaks can be identified

(**Peaks A, B, C and D**) concurrent with the C-H5-C-H6 interaction of the four cytosine peaks. This suggests that the self-complementarity of the duplex has been retained, an observation that is supported by the presence of a single proton walk in this region. It is therefore confirmed that a single *bis*-adduct has been formed.

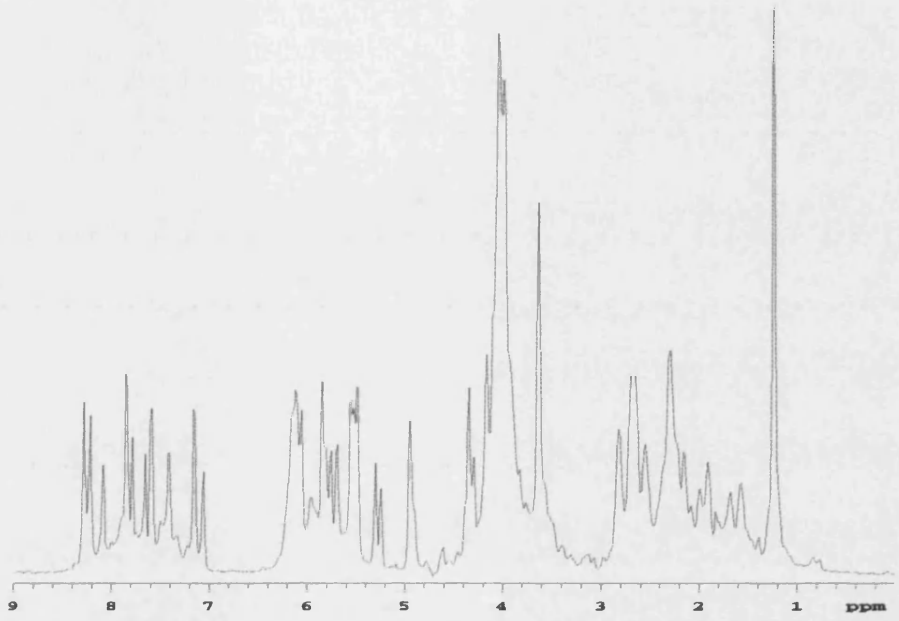


Figure 70: 600 MHz 1D ^{1}H NMR spectrum of the 5'*d*(CICGATCICG)$_2$ duplex

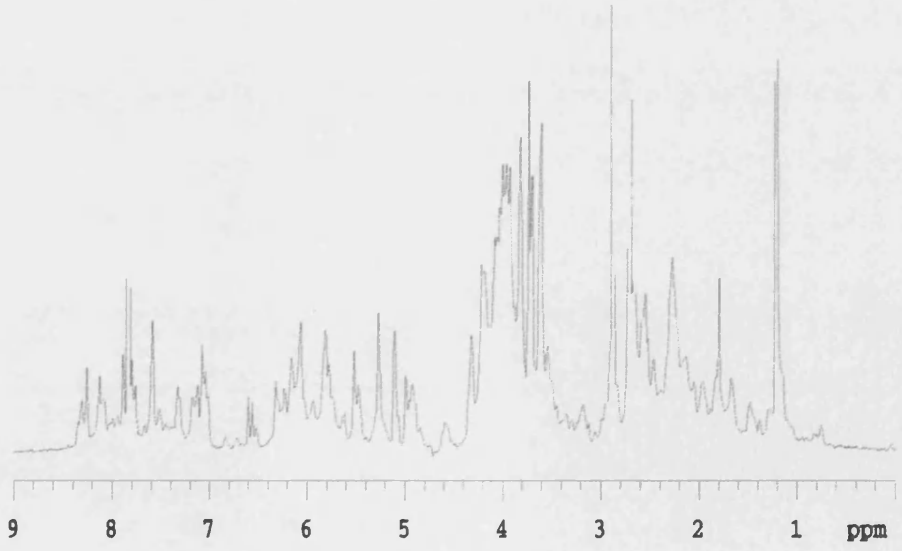


Figure 71: 600 MHz ^{1}H NMR spectrum of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct

Table 12: Chemical Shifts (ppm) of 5'*d*(CICGATCICG)$_2$ and 5'*d*(CICGATCICG)$_2$-SJG-136 adduct DNA Protons

Base	H8/H6			H1'			H2'			H2''			H3'			H4'			H5/CH3/H2		
C1	7.69	7.60	*0.09*	5.56	5.52	*0.04*	1.90	1.82	*0.02*	2.30	2.28	*0.02*	4.63	4.61	*0.02*	3.64	3.63	*0.01*	5.85	5.80	*0.05*
I2	8.39	8.28	*0.11*	6.11	6.15	*0.04*	2.71	2.66	*0.05*	2.82	2.82	*0.00*	4.95	4.91	*0.04*	4.34	4.32	*0.02*	7.92	7.82	*0.10*
C3	7.28	7.34	*0.06*	5.70	5.47	*0.23*	1.68	1.86	*0.18*	2.23	2.27	*0.04*	4.75	4.92	*0.17*	4.03	4.06	*0.03*	5.24	5.26	*0.02*
G4	7.94	7.80	*0.14*	5.49	6.08	*0.06*	2.66	2.66	*0.00*	2.71	2.94	*0.23*	4.93	4.95	*0.02*	4.30	4.06	*0.24*	-	-	-
A5	8.20	8.15	*0.05*	6.14	6.30	*0.16*	2.71	2.46	*0.25*	2.84	2.66	*0.18*	4.90	4.88	*0.02*	4.16	3.94	*0.22*	7.76	8.11	*0.35*
T6	7.17	7.05	*0.12*	5.80	5.71	*0.09*	1.94	1.94	*0.00*	2.37	2.17	*0.20*	4.77	4.55	*0.22*	4.20	3.98	*0.22*	1.23	1.20	*0.02*
C7	7.53	7.14	*0.39*	5.48	5.44	*0.04*	2.00	2.09	*0.09*	2.32	1.65	*0.67*	4.95	4.71	*0.24*	4.05	4.02	*0.03*	5.52	5.26	*0.26*
I8	8.33	8.12	*0.21*	6.08	5.84	*0.24*	2.64	2.57	*0.07*	2.80	2.10	*0.70*	4.94	4.83	*0.11*	4.06	3.82	*0.24*	7.72	7.38	*0.34*
C9	7.28	7.20	*0.08*	5.76	5.77	*0.02*	1.58	1.46	*0.12*	2.15	2.13	*0.02*	4.92	4.71	*0.21*	3.99	3.94	*0.05*	5.30	5.26	*0.04*
G10	7.96	7.90	*0.06*	6.05	6.06	*0.01*	2.28	2.26	*0.02*	2.58	2.55	*0.03*	4.60	4.58	*0.02*	4.10	3.96	*0.14*	-	-	-

Key: Blue = duplex shift (ppm), Black = adduct shift (ppm) Green = +/- Difference (ppm) Underlined = greater than 0.25 ppm

Table 13: Chemical shifts of adduct drug protons

Drug	H1a	H1b	H2a/b	H3a/b	H6	H9	H11	H11a	H12a	H13	H14
Shift (ppm)	2.55	3.21	5.84	4.02/4.06	7.05	6.53	5.64	3.92	4.24	2.28	3.83

Table 14: Comparison of chemical shifts (ppm) 5'*d*(CICGATCICG)$_2$-SJG-136 and 5'*d*(CICGATCICG)$_2$-DSB-120 adducts

	H8/H6			H1'			H2'			H2''			H3'			H4'			H5/CH$_3$/H2		
C1	7.60	7.72	*0.12*	5.52	5.74	*0.22*	1.82	1.92	*0.10*	2.28	2.38	*0.10*	4.61	4.62	*0.01*	3.63	3.98	*0.35*	5.80	5.93	*0.13*
I2	8.28	8.26	*0.02*	6.15	6.21	*0.06*	2.66	2.81	*0.15*	2.82	2.81	*0.01*	4.91	4.94	*0.03*	4.32	4.37	*0.05*	7.82	7.93	*0.11*
C3	7.34	7.50	*0.16*	5.47	5.43	*0.04*	1.86	2.05	*0.19*	2.27	2.30	*0.03*	4.92	4.82	*0.10*	4.06	4.10	*0.04*	5.26	5.52	*0.26*
G4	7.80	7.92	*0.12*	6.08	6.16	*0.08*	2.66	2.71	*0.05*	2.94	3.05	*0.11*	4.95	5.06	*0.11*	4.06	4.12	*0.06*	-	-	
A5	8.15	8.26	*0.11*	6.30	6.44	*0.14*	2.46	2.60	*0.14*	2.66	2.75	*0.09*	4.88	4.98	*0.10*	3.94	4.15	*0.21*	8.11	8.16	*0.05*
T6	7.05	7.07	*0.02*	5.71	5.78	*0.07*	1.94	2.03	*0.09*	2.17	2.19	*0.02*	4.55	4.62	*0.07*	3.98	2.61	*1.37*	1.20	1.27	*0.07*
C7	7.14	7.17	*0.03*	5.44	5.48	*0.04*	1.65	1.65	*0.00*	2.09	2.15	*0.06*	4.71	4.72	*0.01*	4.02	4.07	*0.05*	5.26	5.31	*0.05*
I8	8.12	8.05	*0.07*	5.84	5.86	*0.02*	2.57	2.73	*0.16*	2.10	2.56	*0.46*	4.83	4.89	*0.06*	3.82	3.79	*0.03*	7.38	7.54	*0.16*
C9	7.20	7.44	*0.24*	5.77	5.83	*0.06*	1.46	1.82	*0.36*	2.13	2.34	*0.21*	4.71	4.72	*0.01*	3.94	4.08	*0.14*	5.26	5.52	*0.26*
G10	7.90	7.89	*0.01*	6.06	6.12	*0.06*	2.55	2.64	*0.09*	2.26	2.34	*0.08*	4.58	4.65	*0.07*	3.96	4.15	*0.19*	-	-	

Black = SJG-136 adduct (ppm), Red = DSB-120 adduct (ppm),[181] Green = difference (+/-)(ppm) Underlined = greater than 0.20 ppm

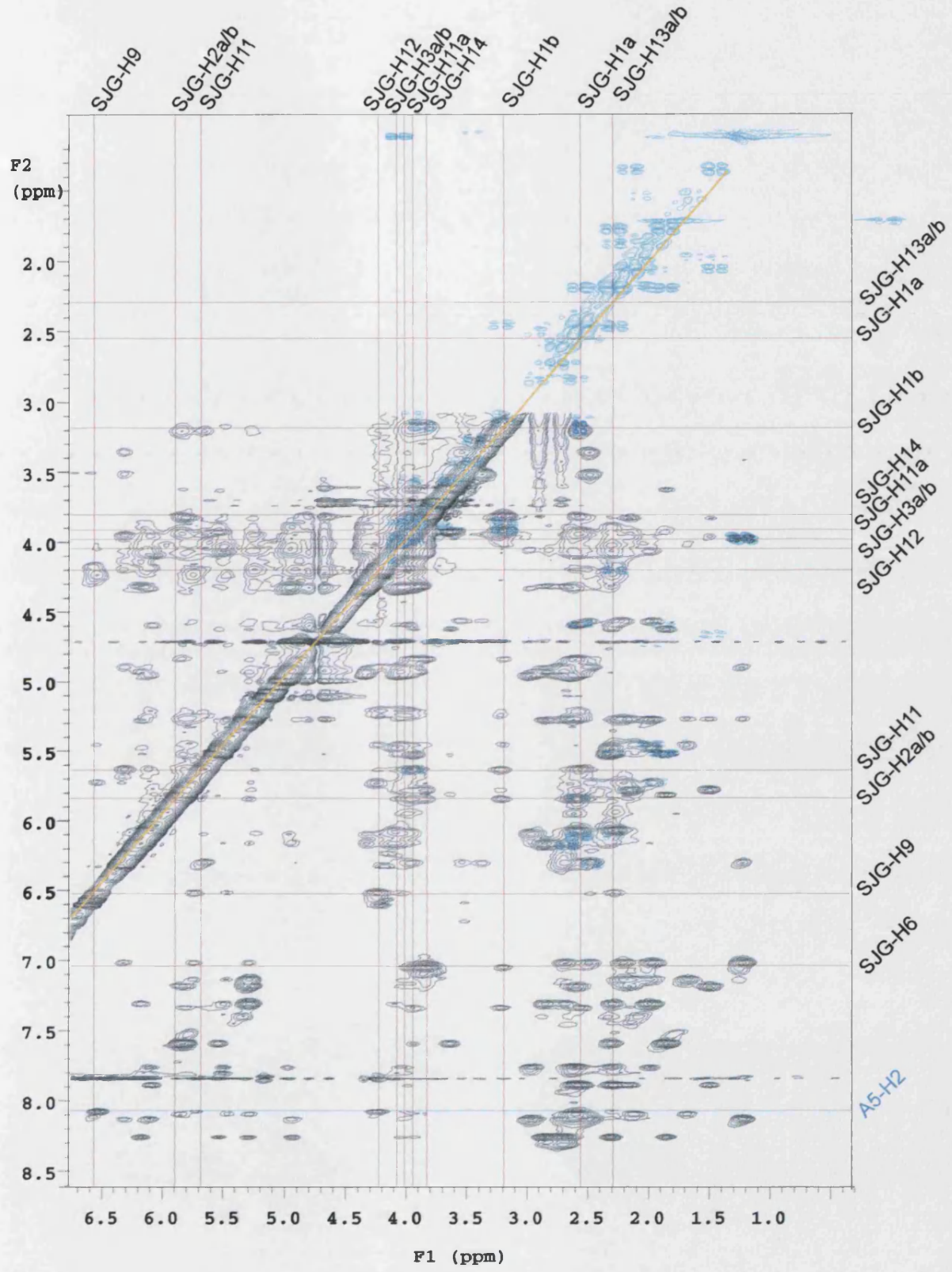


Figure72: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct Black= NOESY, Blue = COSY, NOESY mixing time 200 ms. Complete drug assignments are shown

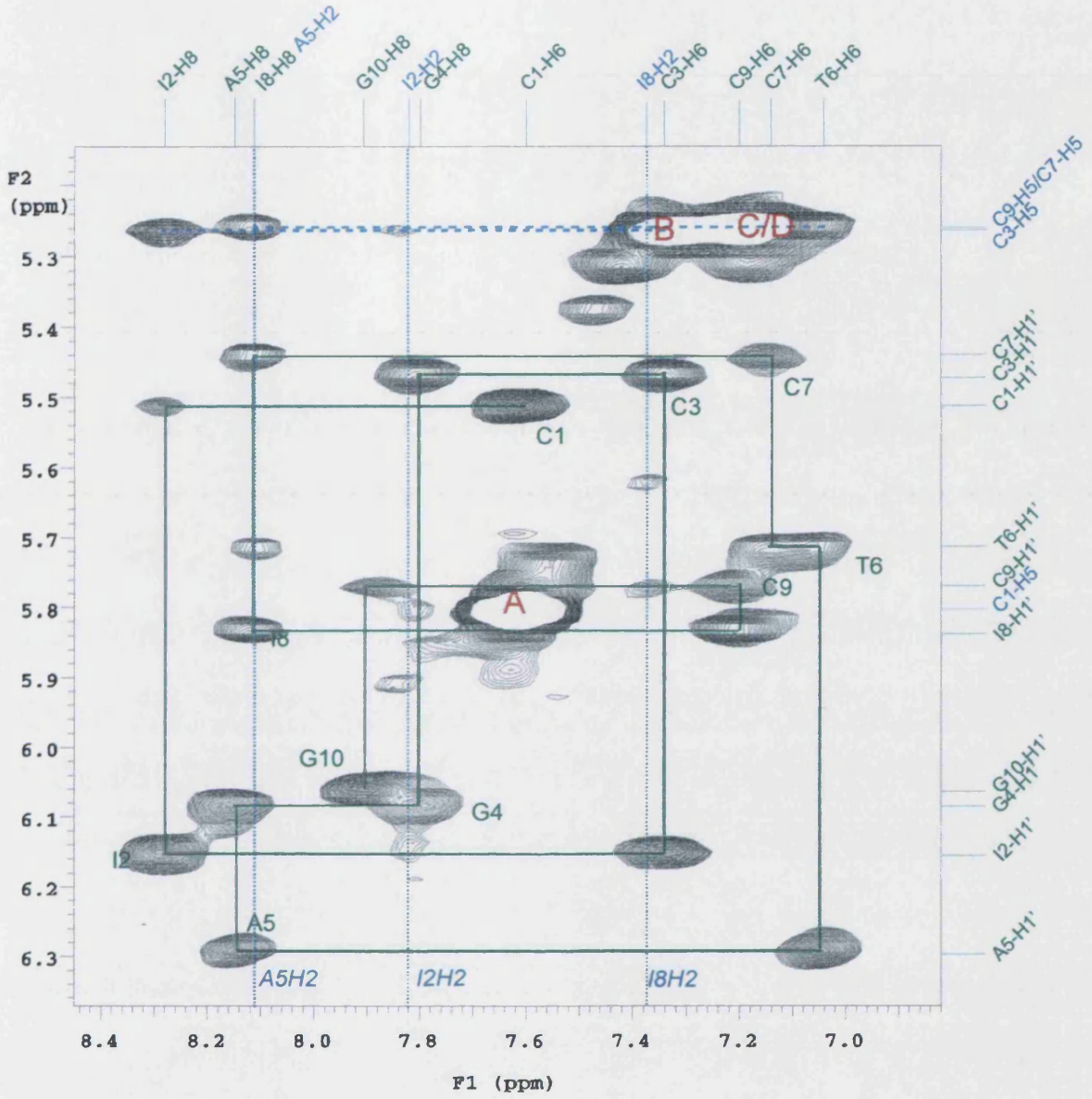


Figure 73: 600 MHz NOESY spectrum of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct Mixing time = 200 ms, H6/H8-H1' region. A, B, C and D show C-H5 to C-H6 cross-peaks

3.3: Identification of the Covalent Linkage Site

PBD ligands are known to react with the *exo*-cyclic NH_2 group of a guanine base *via* either direct attack on the imine species or by an SN2 type mechanism as shown in **Figure 74**[40,41] and therefore only one reaction site is available to this drug on each strand of DNA.

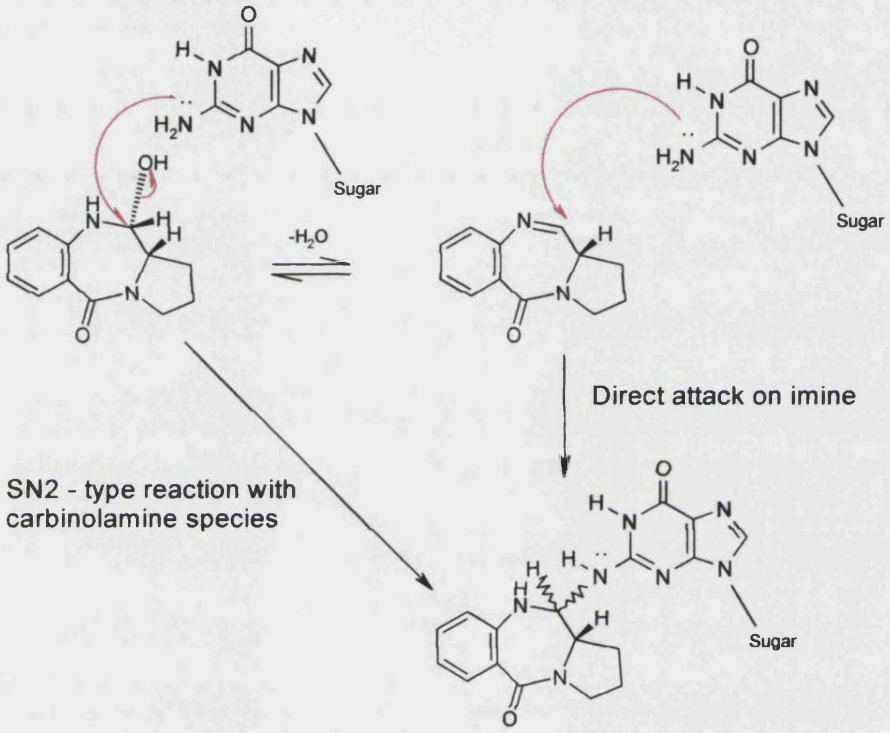


Figure 74: General reaction of a PBD with a guanine base. An equilibrium exists between the imine and carbinolamine species, resulting in possible direct imine attack or SN2 type reactions

Connectivities between SJG-136 and the duplex are listed in **Table 15**. These, along with the large changes in shift value throughout the central bases of the adduct when compared with the duplex (see **Table 12**) confirm the presence of the drug in the minor groove as depicted in **Figure 75.**

	H1a	H1b	H2a/b	H3a/b	H6	H9	H11	H11a	H12a/b	H13a/b	H14
G4-H4'			W								
G4-H1'							M				
A5-H2						S			M		
A5-H1'				W			M				
A5-H4'			M(o)	B			B				
A5-H3'				M							
T6-H4'					W/M				B		B
T6-H5'					W						
T6-H5''					W						
T6-H1'						M			M	M(o)	
C7-H1'		W				W		W(o)	M/W	M(o)	W
C7-H4'						W			B	S(o)	B
I8-H3'	M(o)	M/W									
I8-H2	M	M	W(o)				W	M			
I8-H4'				B	B			B			
I8-H1'				W(o)				M(o)			
C9-H4'			M(o)	B				B			
C9-H1'				W							

Table 15: NOE connectivities between SJG-136 protons and the 5'*d*(CICGATCICG)$_2$ duplex. S = *strong*, M = *medium*, W = *weak*, (o) = *overlaid*, B = *buried*

Furthermore similar shift changes have been previously reported for the reaction of DSB-120 (**Table 14**) with an identical duplex sequence.[181] In light of this evidence the covalent linkage site is identified as G4, supported by large changes in the chemical shift of the opposite base (C7) as well as I8.

3.4: Stereochemistry and Orientation

As discussed in **Chapter 1**, it is possible in theory for either 'S' or 'R' stereochemistry to exist at the covalent linkage site C11 in PBD-DNA adducts (See **Figure 76**).

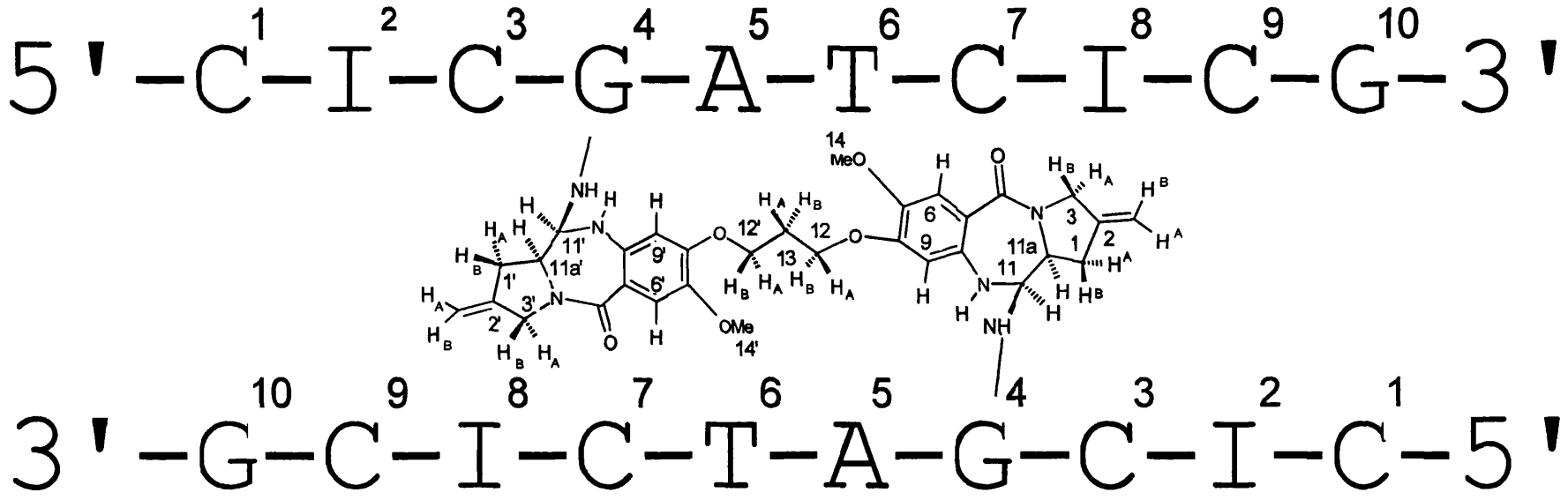


Figure 75: Representation of the 5'*d*(CICGATCICG)$_2$-SJG-136 inter-strand adduct, showing DNA and drug numbering systems

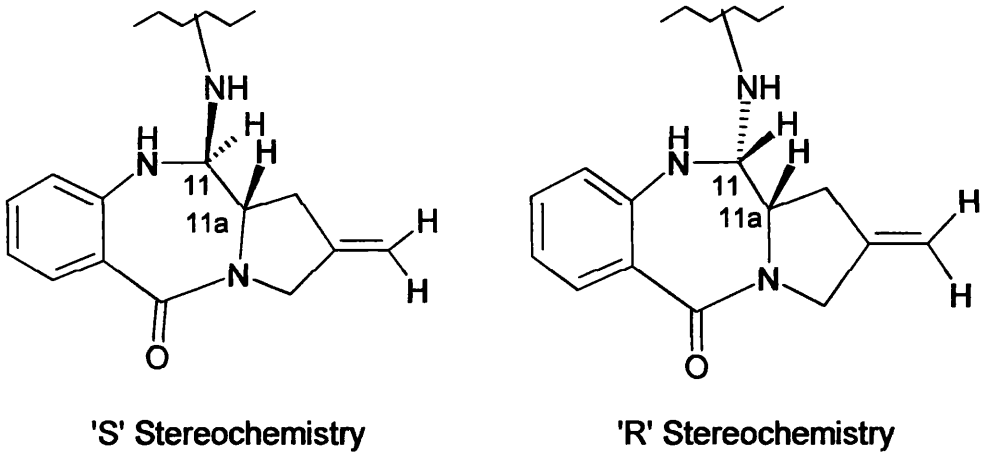


Figure 76: R and S stereochemistry in the PBD subunit

Previous studies have shown that, for the tomamycin (**Figure 69**) *d*(ATGCAT)$_2$ adduct the PBD-11'R' and PBD-11'S' stereoisomers exist in approximately equal proportions.[194] However molecular mechanics calculations predict that the lowest energy conformation is an 11'S' isomer[42,43,44] and other studies have shown the 'S' isomer to be predominant for this class of drugs.[181,195] The proton attached at the PBD-C11 covalent linkage site has proved to be a useful diagnostic probe for determination of the stereochemistry at this site. It has been stated that in an adduct with 'S' stereochemistry at the PBD-C11 point PBD-H11 will be directed towards the 3' side of the covalently modified guanine, whereas for an 'R' isomer the situation is reversed with PBD-H11 closer in space to the 5' side.[181] A pictorial representation of this can be seen in **Figure 77**. In the case of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct a strong NOE between SJG-H11 and A5-H1' (**Figure 78**) confirms that the 'S' isomer is present. This is supported by the absence of any cross-peak between SJG-H11 and the cytosine to the 5' side of G4 (C3). The orientation of the aromatic rings in the SJG-136 adduct is fixed in a head to head orientation by the methylene linker unit.

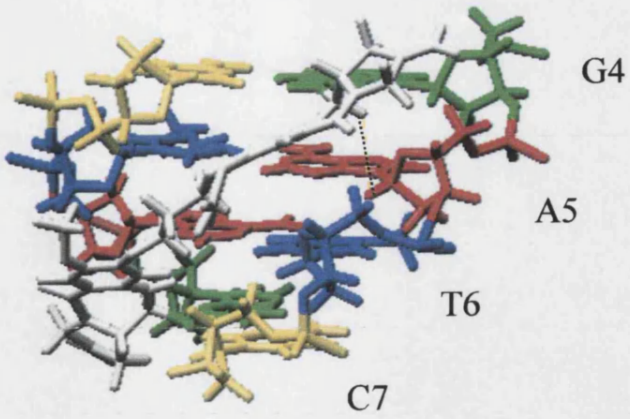


Figure 77: Representation of the central region of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, with 11'S' stereochemistry, showing the proximity of A5-H1' to SJG-H11

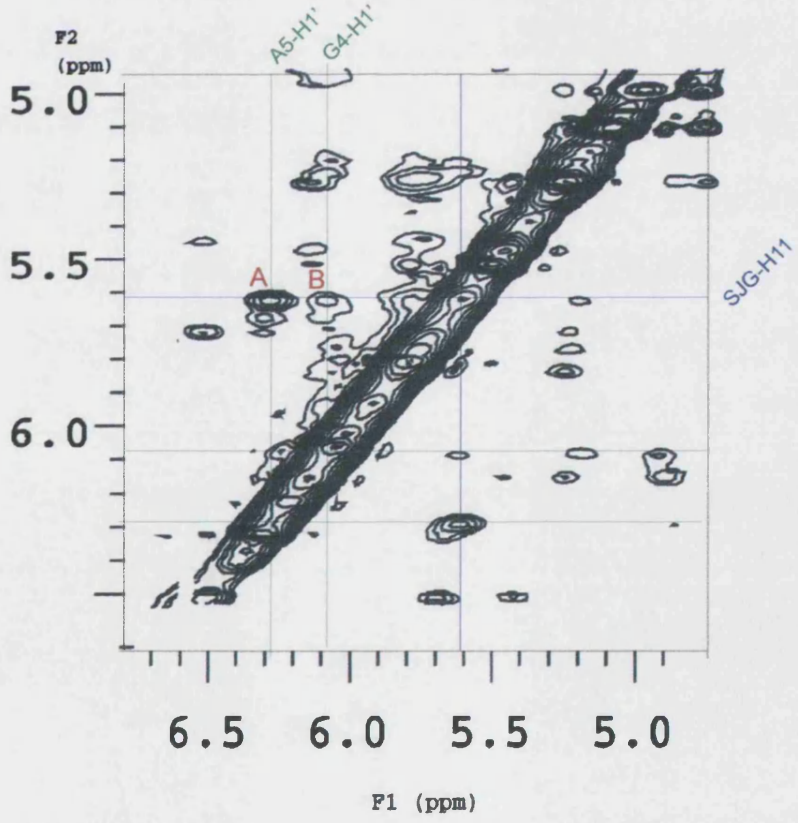


Figure 78: 600 MHz NOESY expansion showing NOEs between SJG-H11 and A5-H1' & G4-H1', mixing time = 200ms

3.5: Nucleic Acid Proton Chemical Shifts

A comparison of the 5'*d*(CICGATCICG)$_2$-DSB-120 adduct[181] with the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct shows a very similar two dimensional NOESY spectrum (see **Table 14**). However, some marked differences were observed. Although the chemical shift values for the two adducts were generally similar there was a marked up-field shift in many of the DNA-C9 base protons in the SJG-136 adduct when compared with the DSB-120 adduct. This, along with changes in the relative shift of DNA-C1 and DNA-I8 protons can be attributed to the shielding effect of the introduction of the ethylene group.

3.5.1: Relative chemical shifts of the DNA-H2' and H2'' protons

As discussed in **Chapter 2**, chemical shifts of H2' and H2'' proton resonances are usually seen between 1.5 ppm and 3.0 ppm. The H1'-H2' coupling is invariably more intense than the H1'-H2'' as a result of the greater inter-proton distance in the latter case and typically the H2' protons resonate up-field of the H2'' protons within the same nucleotide sequence.[196] For the I8 nucleotide in the SJG-136 adduct however, there is a reversal of the general pattern and the H1'-H2'' resonance is found up-field of the H1'-H2'. This feature was also noted in the study of the DSB-120 adduct and is indicative of a conformational change within the internal nucleotide. It suggests that, as with DSB-120, SJG-136 induces an additional perturbation in the structure of the C7 nucleotide facing the covalently modified G4 in the minor groove.

3.5.2: Chemical shift changes in the 5'd(CICGATCICG)$_2$-SJG-136 adduct relative to duplex DNA

Chemical shifts for oligonucleotide proton resonance signals of the SJG-136 adduct relative to the duplex are shown in **Table 12.** There are up-field chemical shifts of 0.39 ppm and 0.26 ppm for C7-H6 and C7-H5 respectively, which mirror shift changes seen when the same duplex sequence was reacted with DSB-120. In addition large shift changes can be seen for C7-H2'/H2'' and C7-H3' protons. These chemical shift

changes suggest an increase in stacking of the T6-C7 step. However studies of the DSB-120 adduct revealed probable de-stacking of the I2-C3 step, deduced from downfield shifts for C3-H6 and C3-H5. These shifts are absent in the SJG-136 adduct spectrum, suggesting that the aforementioned de-stacking does not occur in this case. The I8-H2 resonance of the SJG-136 adduct undergoes a 0.34 ppm up-field shift. Similar observations have been noted in the relative chemical shifts of I8-H2 when the duplex sequence was reacted with tomamycin. A large up-field shift of the I8-H2 relative to the duplex was attributed to the shielding effect of the ethylidene functionality of the drug in this case.[181] This shift is not seen in the DSB-120 adduct due to the absence of the ethylidene group and in the case of SJG-136 it would appear that the introduction of the ethylene functionality produces a similar effect to the former drug. The 0.34 ppm downfield shift of the A5-H5 resonance has been seen in both the DSB-120 and tomamycin adducts also, and can be attributed to de-shielding effects of the aromatic ring of the covalently attached drug. Other chemical shift changes within the adduct relative to the duplex, as seen with DSB-120, are mainly protons in proximity to the attached drug and most are attributable to drug shielding and de-shielding effects.

The previous studies on DSB-120 and tomamycin highlighted a dramatic difference between the chemical shift changes of T6-H4' in each case. This proton is shielded by the drug aromatic rings in both cases but the resulting up-field shift was dramatically attenuated in the case of DSB-120. It is suggested that the structure of the inter-strand cross-linking *via* tethered DNA reactive units in this case causes the DSB aromatic ring to be partially averted from the wall of the groove, greatly reducing the level of shielding experienced. In the case of SJG-136 only a 0.22 ppm up-field shift is observed compared with 1.65 ppm in the case of DSB-120. This suggests that in this case the aromatic rings of the drug are averted still further, causing the shielding to be much reduced relative to that felt by DSB-120. Molecular modeling of this area is shown in **section 3.7**.

3.6: Intermolecular Drug-DNA Contacts

The proton resonances for SJG-136 have been identified and assigned using analysis of the NOESY and COSY spectra as discussed in **Chapter 2.** Contacts have been found between SJG-136 protons and G4, A5, T6, C7, I8 and C9 bases and are shown in **Table 15** as well as pictorially in **Figure79.** The presence of this network of NOE contacts from SJG-136 to specific DNA protons locates the drug indisputably in the minor groove. As seen in studies of the DSB-120 adduct, SJG-H11 shows a strong cross-peak to A5-H1' and a moderate cross-peak to G4-H1' **(Figure 78).** In the case of DSB-120 further evidence for the averted aromatic ring discussed above was found based on the intensity of the cross-peak between SJG-H11 and the A5-H4' proton. Unfortunately, in this study the cross-peak in question is situated under the large COSY peak arising from the scalar coupling of SJG-H11 to SJG-H11a, so no such conclusions can be drawn. However, the methylene protons of the linker (SJG-H12a, H12b, H13a and H13b) all exhibit expected NOESY connectivities that locate them firmly in the minor groove. (See **Table 15 and Figure 79**). As with DSB-120 the characteristic signal between SJG-H13a/b and C7-H4' although overlaid, can be located as a strong cross-peak, while SJG-H12a and SJG-H12b show connectivities to A5-H2 and T6-H1' **(See Figures 80 & 81**). Interestingly, evidence based on connectivities from the SJG-H1a and SJG-H1b protons of SJG-136 to I8-H2 and I8-H1' suggest that this drug is located more deeply in the minor groove than observed for DSB-120. In the study of the latter drug DSB-H1a and DSB-H1b both show very weak NOESY cross-peaks to I8-H2 and I8-H1' and from these reduced through-space connectivities it was deduced that DSB-120 was immersed rather shallowly in the minor groove. In the SJG-136 adduct of the current study the SJG-H1a and SJG-H1b protons show moderate cross-peaks to I8-H2. Unfortunately, as with the A5-H4' proton, the SJG-H1a and SJG-H1b to I8-H1' resonance is too overlaid to be conclusive. However, the deeper location of SJG-136 in the minor groove is further supported by the probable absence of cross-peaks between SJG-H1a/b and A5-H1'/C7-H4'. Three of these peak areas are overlaid but there is a clear absence of any connectivity between SJG-H1a and C7-H4'. This suggests that the SJG-136 adduct bears a greater resemblance to the tomamycin adduct in this respect than to DSB-120, where a bowing of the ends of the drug away from the floor of the minor groove resulted in weak peaks being observed between the protons discussed above.

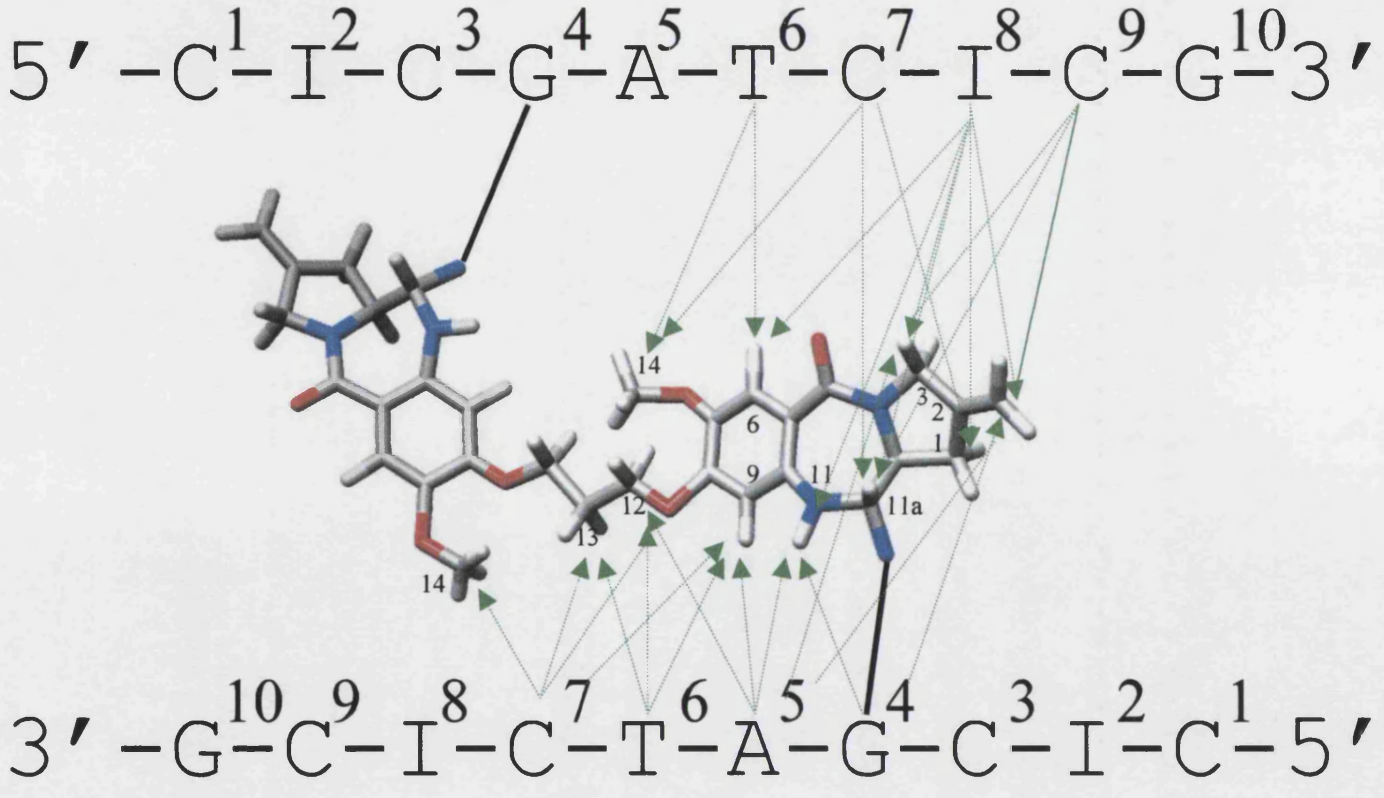


Figure 79: Representation of the 5'*d*(CICGATCICG)$_2$-SJG-136 inter-strand cross-linked adduct, showing NOE connectivities between SJG-136 and the DNA backbone. For increased clarity only one end of the symmetrical molecule is labelled

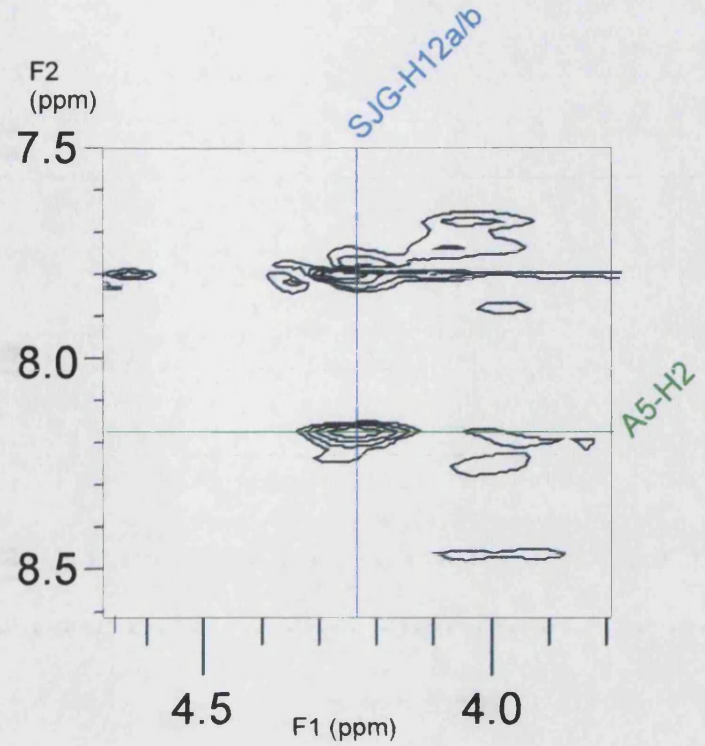


Figure 80: Expansion of 600 MHz NOESY spectrumof the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct showing NOE cross-peak between SJG-H12a/b and A5-H2

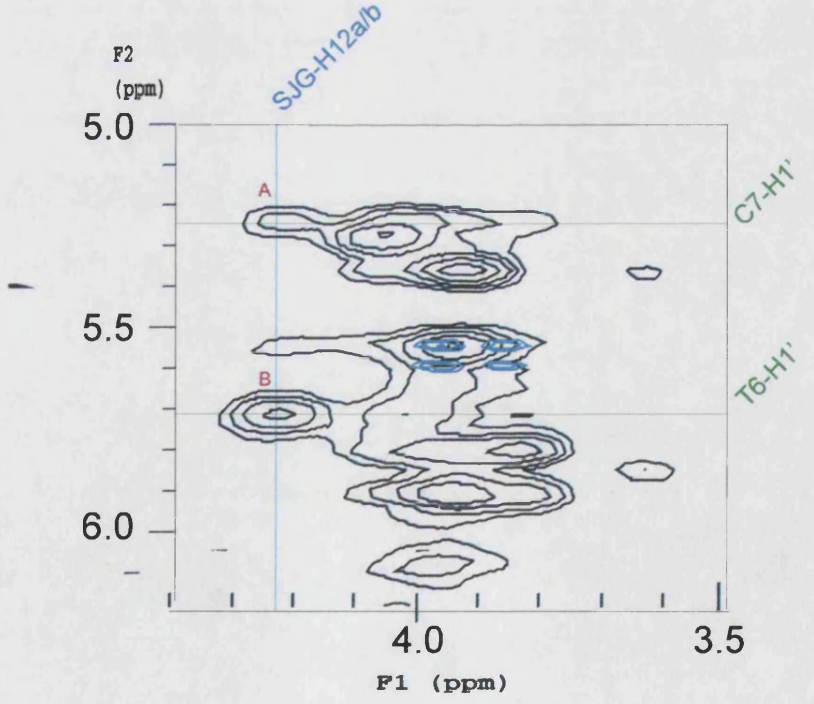


Figure 81: Expansion of 600 MHz NOESY spectrum of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct showing NOE cross-peak between SJG-H12a/b and C7-H1'(A) & T6-H1' (B)

3.7: Molecular modeling of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct

The 5'*d*(CICGATCICG)$_2$-SJG-136 adduct was modelled using the **SYBYL** software suite.[197] This is a general molecular modeling package and was implemented due to reasons of availability and suitability for this type of study. The ligand-DNA complexes were first modelled using the **HYPERCHEM** package.[198] Both DNA and drug were drawn and docked within the package and subjected first to steepest descent energy optimisation and secondly to molecular dynamics *in vacuo* at 300 K. SJG-136 was isolated, minimised and then imported from **HYPERCHEM** into the **SYBYL** suite where it was re-docked onto a DNA duplex produced within **SYBYL**. This procedure ensured correct atom typing/charges within the final model.

The following procedures were then implemented to produce a refined structural model of the adduct.

1. Restraints were applied to the DNA H-bonds of the terminal bases to reduce end fraying during MD calculations.
2. The charges were recalculated for the modified bases and ligand using the Gasteiger Huckel method within the **SYBYL** package. Standard DNA charges were used for the rest of the DNA duplex.
3. The ligand molecule and modified bases were subjected to steepest descent energy minimization using minimization software supplied within the **SYBYL** package.
4. The entire adduct was then subjected to steepest descent energy minimization.
5. The adduct was solvated with 5 layers of Gasteiger Hückel charged water molecules within a droplet. A single sodium counter ion was associated with each end of the DNA backbone phosphates.
6. Molecular mechanics calculations were then performed on the adduct. The temperature was ramped in stages/steps from 0K to 300K over around 50 ps and then held at 300 K for 20 ps. All structure refinement utilised the Tripos force field and **AMBER** mechanics.
7. When a satisfactory model had been obtained steps 4-6 were repeated applying the restraints generated from the 2D NMR datasets (NOESY and COSY). Full restraints can be found in **Appendix I**. The additional restraints initially used

were maintained on the terminal bases to reduce end fraying of the duplex during the final stages of the calculation.

8. The final 10ps of the rMD data (at 300 K) were averaged and the drug-DNA model was subjected to an energy minimization procedure *in vacuo* to produce the final NOE refined molecular model of the adduct.

3.7.1: The importance of solvation and counter ions in molecular mechanics operations

Forces between charged atoms maintain the macromolecular tertiary structure in all nucleic acids. Westhof *et al*[199] stated that β-helical structures were sustained by an equilibrium between several factors. These included electrostatic forces between negatively charged phosphates, hydrophobic interactions between aromatic rings of bases (promotes orderly base pair stacking) and energy of the sugar-phosphate backbone conformation.

Non-covalent bonds within the helix of DNA consist of electrostatic forces, hydrogen bonding and *Van der Waals* forces. These are all influenced greatly by water in all biological systems and Stryer[16] stated that the addition of water could reduce the strength of electrostatic interactions by up to 80 times relative to those experienced by DNA *in vacuo*. Therefore it was deemed important to solvate the drug-DNA adduct in order to give an accurate representation of its structure in biological systems. Hence the inclusion of 5 layers of Gasteiger Hückel charged water molecules all subjected to calculations in the molecular mechanics of the refined model.

3.7.2: The refined molecular model of the 5'd(CICGATCICG)$_2$-SJG-136 adduct

Molecular mechanics and dynamics calculations on the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct were performed *in aquo* using the standard method outlined in this chapter. All figures are from the minimised averaged structure, the water and counter ions were removed for clarity.

Figures 82 - 84 show the *in aquo* computer model of the SJG-136-5'*d*(CICGATCICG)$_2$ DNA duplex. SJG-136 is depicted in white, with the two DNA strands in green and purple. It is clear from this picture that Watson-Crick base pairing has been maintained. This is in agreement with the 2D NOESY NMR spectrum, which suggested the DNA had retained the β-helical structure required for the sequential assignment procedure to be successful. It can be seen that the drug is crescent shaped and fits snugly into the minor groove as expected, with minimal protrusion. This is in analogy to all other PBD drugs studied in this manner to date. **Figure 85** shows the central region of the duplex with SJG-136 in *situ.* The central AT base pairs are shown in red and appear to have become slightly distorted in order to accommodate the drug. This is supported by the very weak A5-H8 to T6-H6 cross-peak visible in the aromatic region of the spectrum. (**See Figure 86**). **Figure 87** shows the T6-H4' proton discussed in section 3.6.2. DNA strands are shown in green and purple with SJG-136 depicted in white. The T6-H4' proton in question, as well as the drug aromatic ring are shown in red. NMR results suggested that the aromatic ring is partially averted from the wall of the minor groove resulting in an attenuation of the shielding received by T6-H4'.

3.8: Conclusions

In summary, a novel 5'*d*(CICGATCICG)$_2$-SJG-136 adduct has been produced and examined by high field NMR spectroscopy. The drug binds covalently in the minor groove to the *exo*-cyclic NH_2 groups of guanines G4 on opposite strands of DNA, forming an inter-strand cross-link with SJG-C11 'S' stereochemistry at both reaction sites.

Self-complementarity is maintained and minimal disruption to the DNA backbone is observed, allowing retention of a β-helical structure and subsequent assignment of all proton resonances using a sequential assignment procedure.

Molecular models of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct have been generated using distance restraint data compiled from the 2D NMR experiments performed on the adduct. These show the drug fitting comfortably into the minor groove, following the contour of B-form DNA as a result of the 'S' stereochemistry seen at the SJG-C11a

chiral centre. Some distortion is evident in the central base pairs as a result of the accommodation of the drug in this fashion.

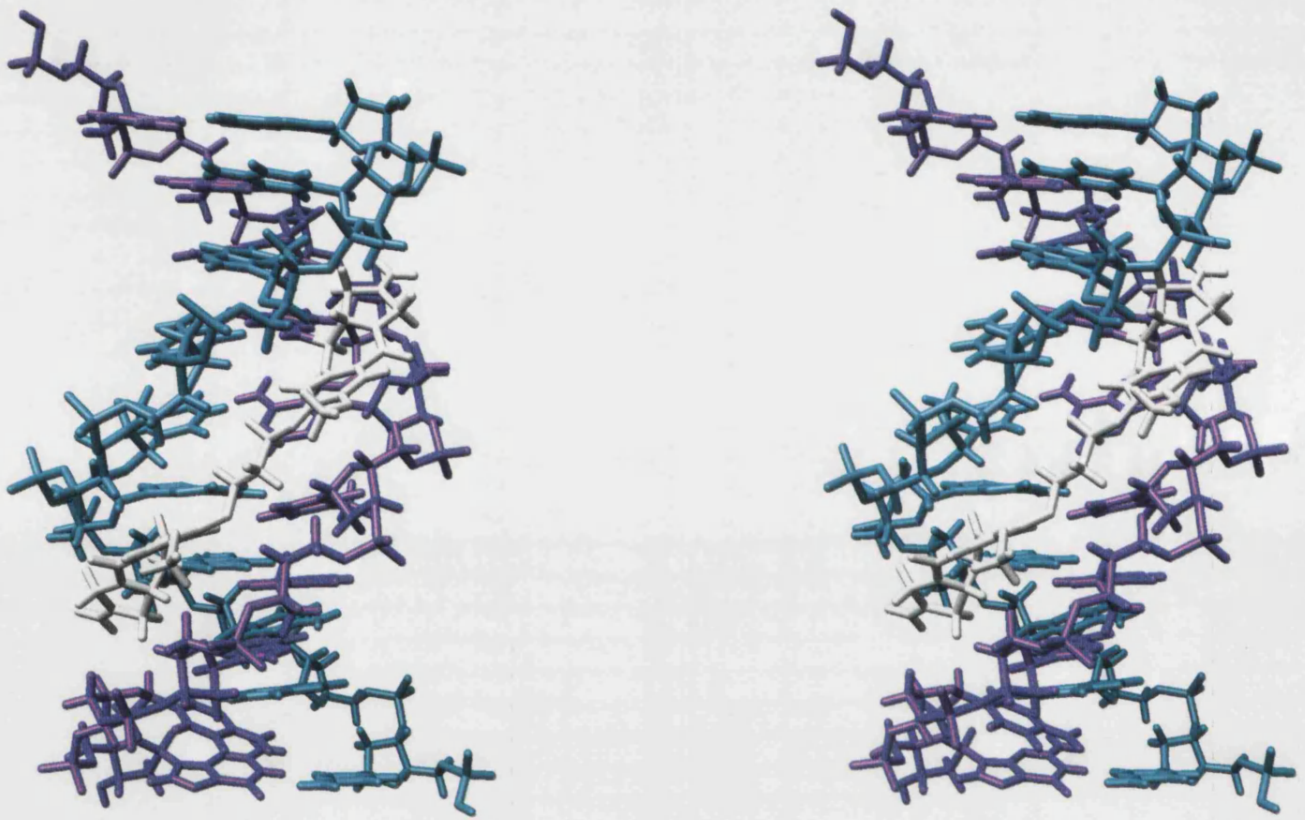


Figure 82: Cross-eye stereo view of 5'*d*(CICGATCICG)$_2$-SJG-136 adduct. DNA strands are shown in green and purple, SJG-136 is depicted in white. Model generated using SYBYL[197] and pictured using UCSF CHIMERA[193]

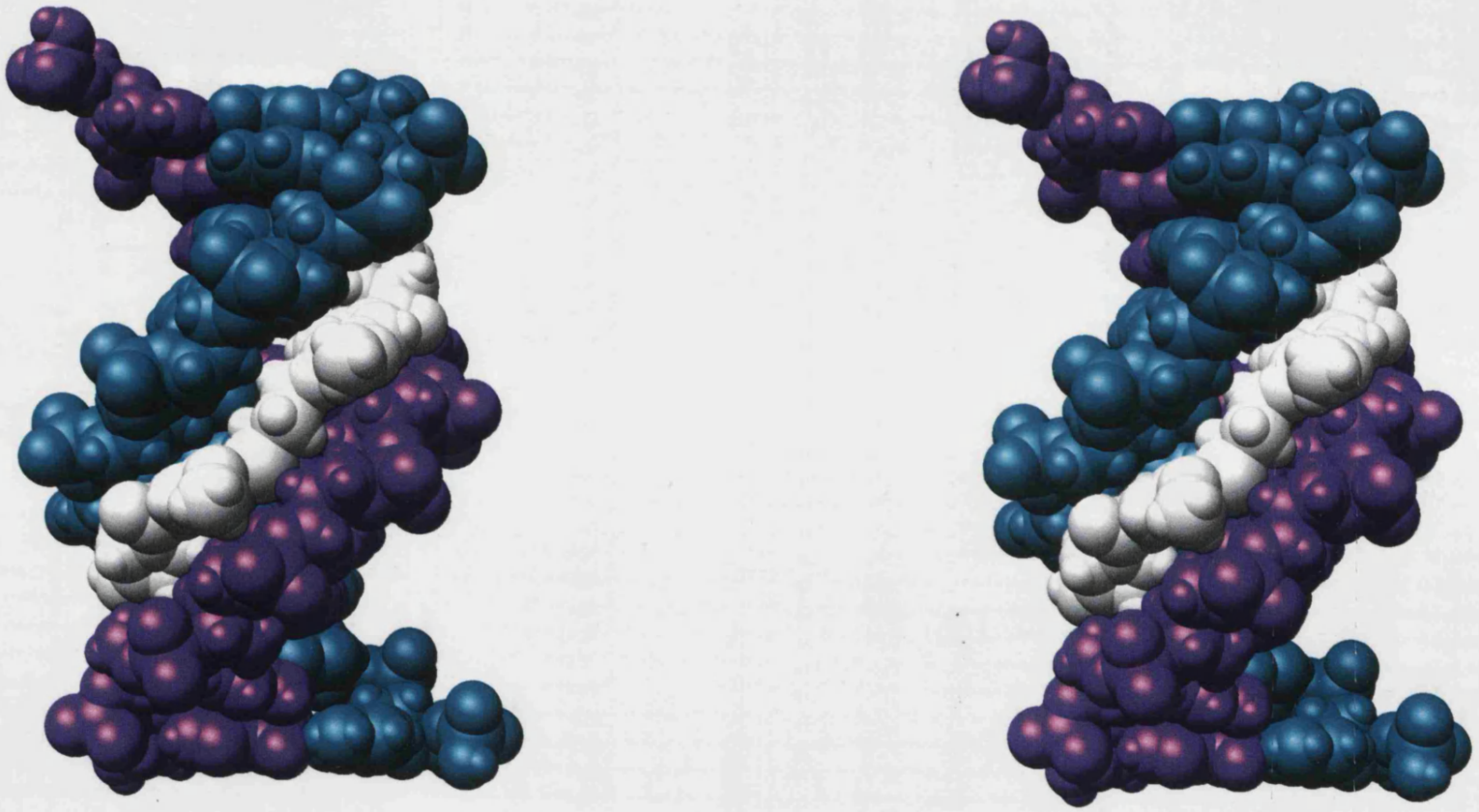


Figure 83: Cross-eye stereo view of 5'*d*(CICGATCICG)$_2$-SJG-136 adduct., showing snug fit of SJG-136 into the minor groove
DNA strands are shown in green and purple, SJG-136 is depicted in white.
Model generated using SYBYL[197] and pictured using CHIMERA[193]

Figure 84: Cross-eye stereo view of 5'*d*(CICGATCICG)$_2$-SJG-136 adduct. View from above showing maintenance of β-helical structure and fit of SJG-136 within the boundaries of the DNA. DNA strands are shown in green and purple, SJG-136 is depicted in white. Model generated using SYBYL[197] and pictured using UCSF CHIMERA[193]

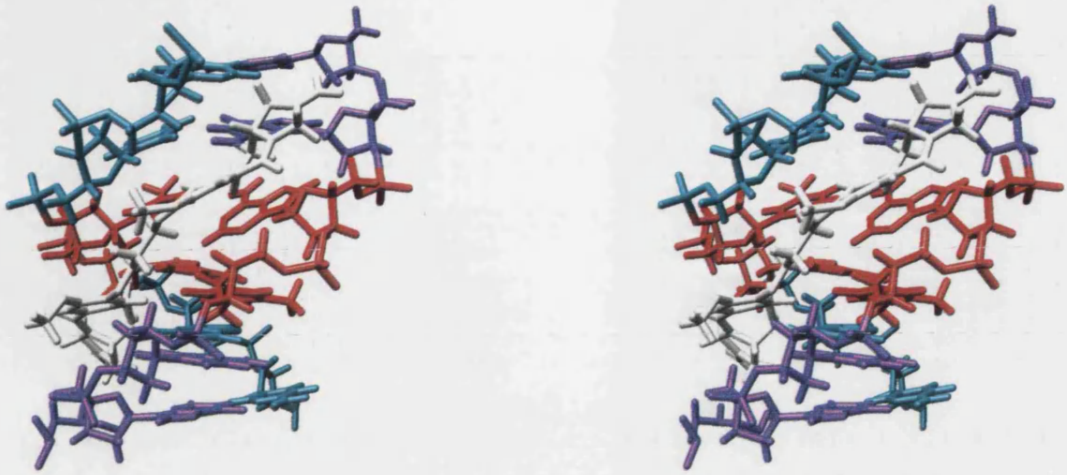


Figure 85: Cross-eye stereo view of 5'*d*(CICGATCICG)$_2$-SJG-136 adduct. Central region only showing distortion of central bases. DNA strands are shown in green and purple, SJG-136 is depicted in white. Central A-T base pairs are shown in red. Model generated using SYBYL[197] and pictured using UCSF CHIMERA[193]

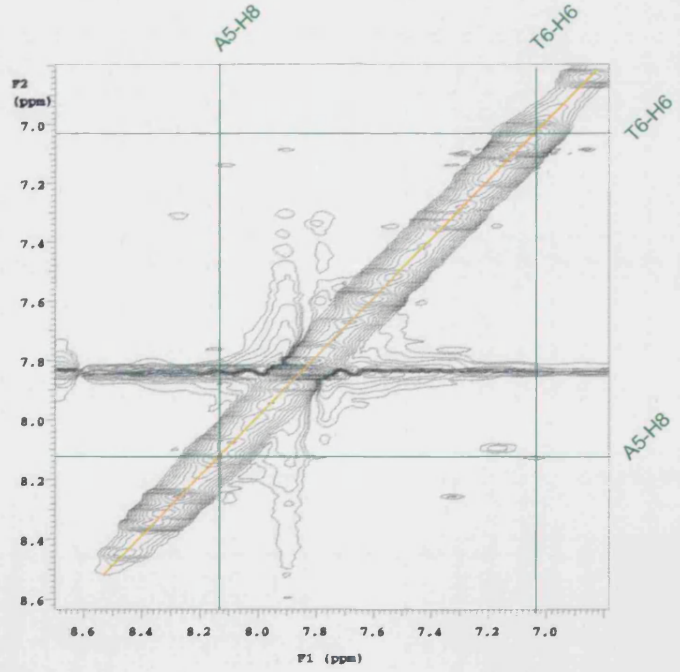


Figure 86: Expansion of the 600MHz 2D NOESY spectrum of 5'*d*(CICGATCICG)$_2$-SJG-136 showing the H8/H6 – H8/H6 region. A very weak cross-peak between A5-H8 and T6-H6 can be observed as a result of the distortion of the central A5 and T6 base pairs

Figure 87: Cross-eye stereo view of 5'*d*(CICGATCICG)$_2$-SJG-136 adduct.
Central region only showing proximity of T6-H4' to the PBD aromatic ring. DNA strands are shown in green and purple, SJG-136 is depicted in white. Aromatic ring and H4' proton are shown in red.
Model generated using SYBYL[197] and pictured using CHIMERA[193]

CHAPTER 4: RESULTS AND DISCUSSION

THE 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 INTRA-STRAND ADDUCT

4.1: Design and Synthesis of Duplex

The 5'*d*(CTCATCAC).(GTGATGAG) duplex was designed to afford no reaction sites on the 'C' strand, while providing two available sites on the 'G' strand in order to investigate the possibility of intra-strand cross-link formation. SJG-136 is known to react *via* direct attack or an SN2 type reaction mechanism through the *exo*-cyclic NH_2 group of a guanine base. This means that the only possible reaction route for the drug with 5'*d*(CTCATCAC).(GTGATGAG) is to react at guanine bases G11 and G14 as shown in **Figure 88.** SJG-136 has been shown to cover a three base pair region hence the DNA bases G9 and G14 are situated too far apart for a cross-linking reaction at these sites to be feasible, while the cytosine bases C3 and C6 contain no amino group at which reaction can occur.

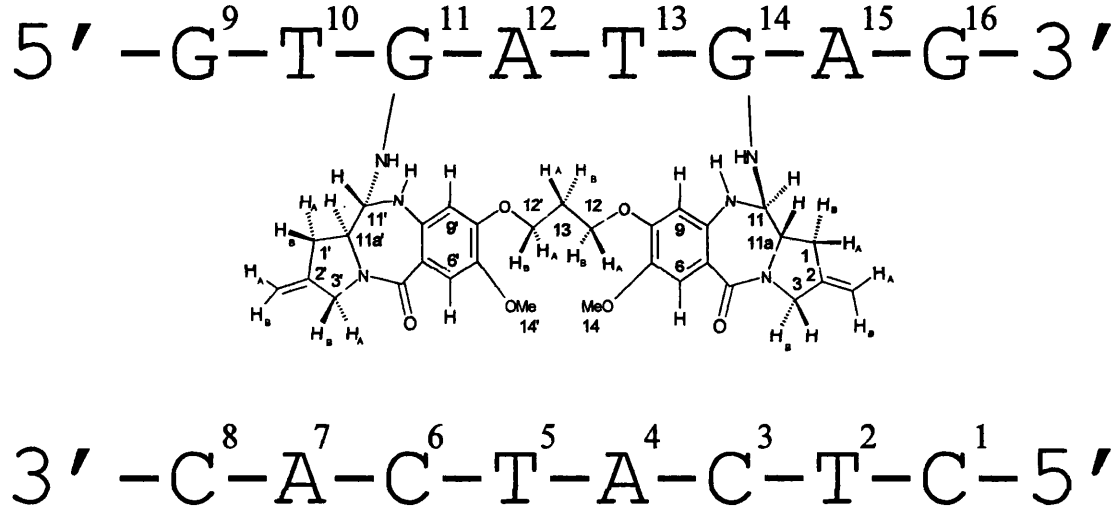


Figure 88: Representation of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand cross-linked adduct showing the drug numbering system

The duplex was synthesised using methods discussed in **Chapter 2.** As for the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, the duplex was allowed to react for 16 hours with 5 mg of SJG-136. From the 1D ^{1}H NMR spectrum it was deduced that the reaction had proceeded to completion and 2D NOESY and COSY experiments were performed on the resulting adduct.

4.2: Confirmation of cross-link formation with two alkylation sites

The one-dimensional ^{1}H-NMR non-exchangeable spectra for both the duplex and the adduct can be seen in **Figures 89 & 90.** Once again the majority of signals are well-resolved - it is clear that the reaction has proceeded to completion, and a new species has been formed. An aromatic peak count is difficult due to the presence of overlaid signals towards the centre of the region, but an increase in the number of peaks can be observed, as would be expected with the addition of SJG-H9 and SJG-H6, both of which may give rise to resonances in the aromatic region. In addition a new resonance is clearly discernable at around 6.4 ppm, while the methyl peak present at 1.2 ppm in the duplex spectrum is no longer visible – suggesting that there is no remaining duplex DNA in the solution. A count of methyl peaks between 1-2 ppm yields a total of six as expected for a non-self-complementary DNA-SJG-136 adduct (four thymine methyl peaks plus two SJG-methyl peaks) in comparison with the duplex spectrum, where only four methyls corresponding to the four thymine bases can be discerned.

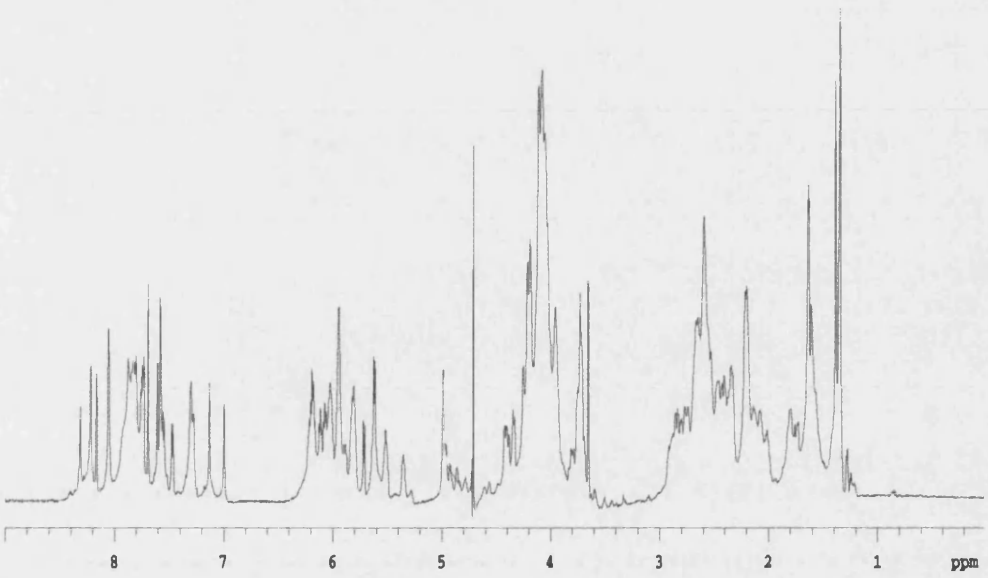


Figure 89: 400 MHz ^{1}H NMR spectrum of the 5'*d*(CTCATCAC).(GTGATGAG) duplex

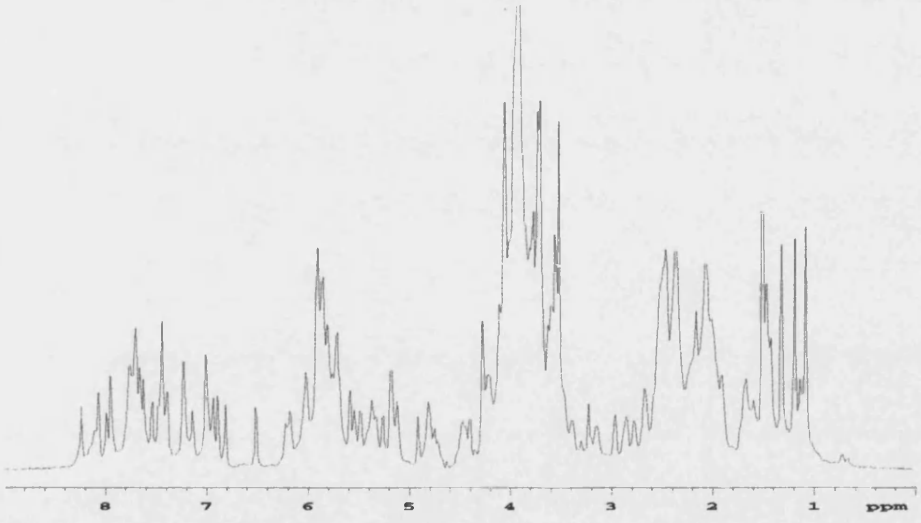


Figure 90: 400 MHz ^{1}H NMR spectrum of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct

For proton assignment of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct once again COSY and NOESY NMR techniques were employed. An expansion of the H8/H6-H1' region of the duplex and adduct NOESY spectra can be seen in **Figure 91**. It is clear when the two spectra are superimposed that there are no duplex resonances remaining in the adduct spectrum, confirming that the reaction has proceeded to completion as suggested by examination of the 1D spectra.

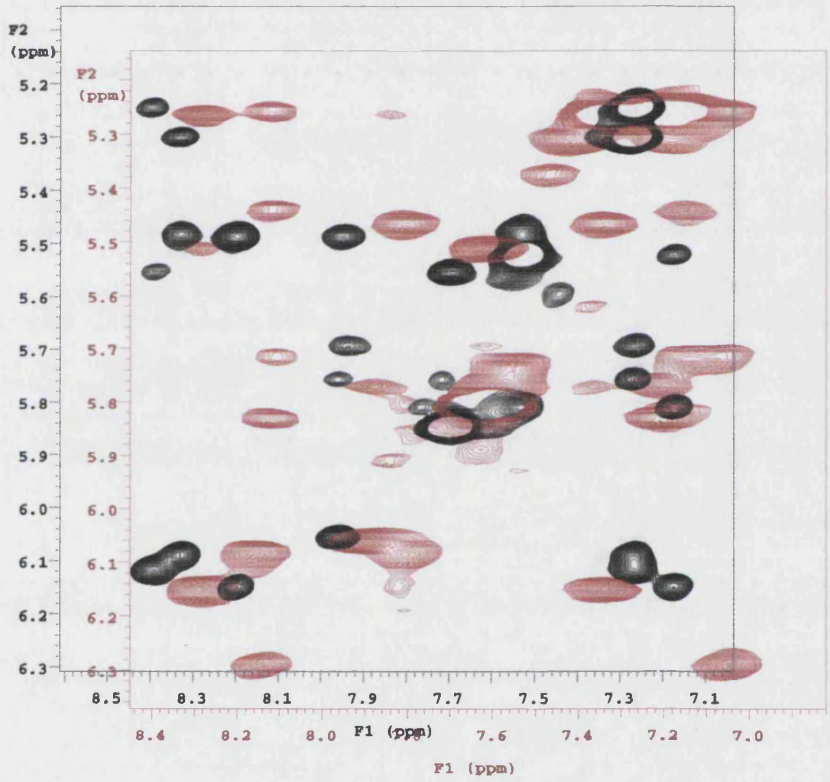


Figure 91: 600 MHz NOESY spectra of the 5'*d*(CTCATCAC).(GTGATGAG) duplex (black) and the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct (red), H6/H8 to H1' region

Four strong COSY peaks can be identified in the H6/H8-H1' region, confirming the presence of only one adduct. The region was fully assigned using the sequential assignment procedure outlined in **Chapter 2**, and can be seen in **Figure 92**. The presence of two complete 'walks' in this region confirms that the DNA strand is non-

self-complementary as expected, and has retained its B-form character. In addition, two complete drug proton walks are possible across the whole spectrum, showing that the drug has reacted at both ends forming a non-self-complementary *bis*-adduct. Complete chemical shift assignments for the nucleotide and drug protons in the duplex and the adduct are listed in **Tables 16 - 18.**

4.3: Identification of the Covalent Linkage Site

As previously discussed, PBD ligands are known to react with the *exo*-cyclic NH_2 group of a guanine base as shown in **Figure 74 (Chapter 3)**. The 5'*d*(CTCATCAC).(GTGATGAG) duplex as discussed previously was designed to afford no reaction sites on the 'C' strand, with two available sites on the 'G' strand in order to promote an intra-strand linkage.

As with the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct, large changes in chemical shift at G14-H1' and G11-H1' (0.85 ppm and 0.60 ppm respectively) as well as their Watson-Crick paired cytosines, help to confirm these bases as the sites of covalent modification. Large changes in chemical shift are also seen in the central protons A4-H1' and T5-H1' ('C' strand) and T13-H1' and A12-H1' ('G' strand). In addition, relatively large changes (> 0.20 ppm) are seen for several of the 'G' strand H2'' protons (G9-H2'', T10-H2'' and T13-H2''). This is suggestive of a greater association of the drug with the modified strand, and is not unexpected.

Connectivities between SJG-136 and the duplex were numerous and are listed in **Tables 19 & 20 (section 4.6)** A pictorial representation is shown in **Figure 96.** These connectivities, in conjunction once again with large changes in the chemical shift value throughout the central bases of the adduct when compared with the duplex, place the drug unequivocally within the minor groove of the DNA duplex. This, along with confirmation of the covalent linkage sites, confirms the presence of the first **intra-strand** linkage seen in these complexes.

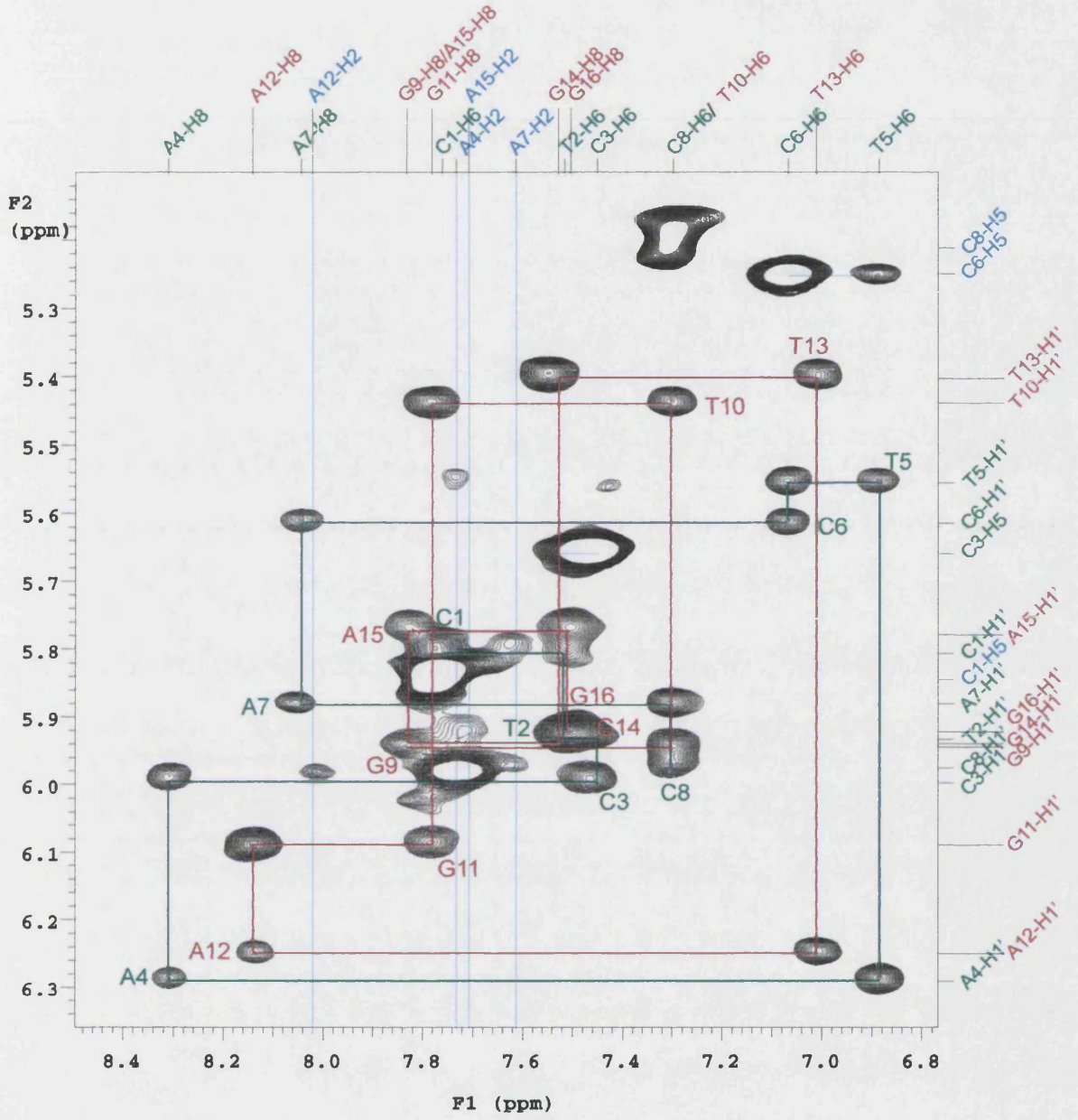


Figure 92: 600 MHz NOESY spectrum of 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136, 200ms mixing time, H6/H8 to H1' region

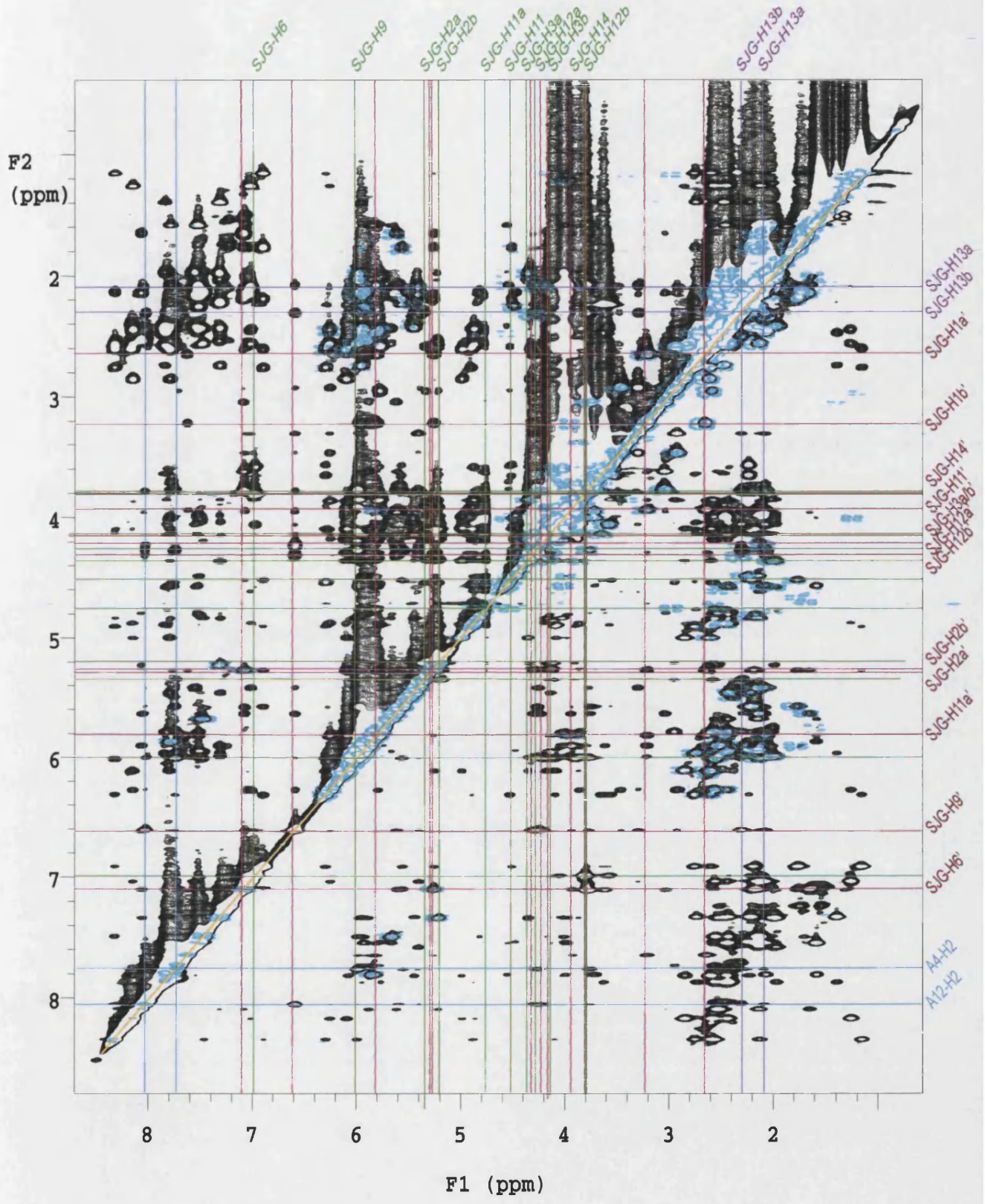


Figure 93: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct. Black = NOESY, Blue = COSY. NOESY mixing time 200 ms

	H8/H6			CH_3/H2/H5			H1'			H2'			H2''			H3'			H4'		
C1	7.70	7.76	*0.06*	5.76	5.84	*0.08*	5.67	5.80	*0.13*	2.04	2.08	*0.02*	2.38	2.44	*0.06*	4.46	4.55	*0.09*	3.90	3.98	*0.08*
T2	7.48	7.52	*0.04*	1.50	1.58	*0.08*	5.96	5.93	*0.03*	2.04	2.15	*0.11*	2.40	2.44	*0.04*	4.70	4.79	*0.09*	4.05	4.13	*0.17*
C3	7.42	7.45	*0.05*	5.56	5.65	*0.09*	5.34	5.99	*0.65*	1.98	2.12	*0.14*	2.26	2.55	*0.29*	4.40	4.82	*0.42*	3.96	4.18	*0.22*
A4	8.18	8.32	*0.14*	7.45	7.73	*0.28*	6.05	6.29	*0.24*	2.47	2.75	*0.28*	2.74	2.60	*0.14*	4.83	4.89	*0.06*	3.96	3.89	*0.07*
T5	7.00	6.89	*0.11*	1.24	1.15	*0.09*	5.66	5.55	*0.11*	1.84	2.18	*0.34*	2.24	1.76	*0.48*	4.64	4.51	*0.13*	3.96	3.56	*0.40*
C6	7.35	7.07	*0.28*	5.46	5.25	*0.21*	5.37	5.61	*0.24*	1.86	2.12	*0.26*	2.18	1.64	*0.54*	4.64	4.71*	*0.07*	3.90	3.94	*0.04*
A7	8.10	8.04	*0.06*	7.67	7.61	*0.06*	6.03	5.88	*0.13*	2.46	2.53	*0.07*	2.66	2.42	*0.24*	4.83	4.78	*0.05*	3.96	3.62	*0.34*
C8	7.19	7.30	*0.11*	5.20	5.21	*0.01*	5.86	5.97	*0.11*	1.90	1.97	*0.07*	2.64	2.02	*0.62*	4.60	4.71*	*0.11*	4.27	4.04	*0.23*
G9	7.77	7.83	*0.07*	-	-	-	5.80	5.94	*0.14*	2.05	2.47	*0.42*	2.60	2.72	*0.12*	4.77	4.71*	*0.06*	3.90	4.13	*0.23*
T10	7.19	7.30	*0.11*	1.20	1.39	*0.19*	5.64	5.44	*0.20*	1.98	2.42	*0.44*	2.32	2.22	*0.10*	4.70	4.82	*0.12*	4.03	4.13	*0.10*
G11	7.73	7.78	*0.05*	-	-	-	5.48	6.08	*0.60*	2.51	2.84	*0.33*	2.62	2.60	*0.02*	4.83	4.97	*0.14*	4.16	4.35	*0.19*
A12	8.04	8.14	*0.10*	7.60	8.02	*0.42*	6.01	6.24	*0.23*	2.41	2.55	*0.14*	2.70	2.44	*0.26*	4.76	4.86	*0.10*	4.26	3.86	*0.40*
T13	6.86	7.01	*0.15*	1.19	1.26	*0.07*	5.46	5.40	*0.06*	1.68	2.15	*0.47*	2.04	2.10	*0.06*	4.83	4.60	*0.23*	3.89	3.45	*0.44*
G14	7.66	7.52	*0.14*	-	-	-	5.19	5.94	*0.85*	2.32	2.45	*0.13*	2.44	2.57	*0.13*	4.62	4.55	*0.07*	4.10	4.05	*0.05*
A15	7.92	7.83	*0.09*	7.60	7.71	*0.11*	5.92	5.77	*0.15*	2.44	2.44	*0.00*	2.68	2.55	*0.13*	4.76	4.87	*0.11*	4.23	4.05	*0.18*
G16	7.44	7.51	*0.07*	-	-	-	5.78	5.92	*0.13*	2.16	2.08	*0.08*	2.38	2.55	*0.17*	4.82	4.47	*0.37*	**	3.98	-

Table 16: Complete chemical shift assignments for the 5'*d*(CTCATCAC).(GTGATGAG) duplex and 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct. Black = duplex, Red = adduct, Green = difference (greater than 0.2 ppm underlined). * = unfound, assigned as under residual water peak at 4.71 ppm, ** = unfound.

H1a	H1b	H2a	H2b	H3a	H3b	H6	H9	H11	H11a	H12a	H12b	H13b	H13a	H14
2.95	2.58	5.32	5.17	4.32	4.11	6.95	5.97	4.52	4.74	4.25	3.76	2.30	2.07	3.80

Table 17: Chemical shift assignments for SJG-136 protons in the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, G-3' end.

H1a'	H1b'	H2a'	H2b'	H3a/b'	H6'	H9'	H11'	H11a'	H12a'	H12b'	H14'
3.22	2.64	5.24	5.22	4.09/4.14	7.06	6.58	5.76	3.91	4.20	4.30	3.78

Table 18: Chemical shift assignments for SJG-136 protons in the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, G-5' end.

4.4: Stereochemistry at alkylation sites

As discussed in **Chapter 3**, the proton attached at the C11 covalent linkage site has proved to be a useful diagnostic probe for the determination of 'R' or 'S' stereochemistry at this site.[181] In an adduct with 'S' stereochemistry at SJG-C11 the SJG-H11 proton will be directed towards the 3' side of the covalently modified guanine, with an 'R' isomer leading to a reversed situation with the SJG-H11 proton closer in space to the 5' side (**See Figure 95**). In the case of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct the 5' (G11) end of the duplex shows a strong NOE between SJG-H11 and A12-H1'. (**see Figure 94**). This confirms that 'S' stereochemistry at SJG-C11 is present at this end of the molecule. This is not unexpected as both the previous adduct (**Chapter 3**) and the 5'*d*(CICGATCICG)$_2$-DSB-120 adduct[181] displayed 'S' stereochemistry at this site. At the 3' end of the duplex the PBD subunit is a mirror image of that at the 5' end. A SJG-C11 centre with 'S' stereochemistry will therefore point toward the 3' neighbour of G14's Watson-Crick base paired cytosine on the opposite strand (A4). 'R' stereochemistry results in a connectivity with G14's own 3' neighbour, A15. When the 2D NOESY spectrum for the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct is examined there is no connectivity between SJG-H11 and A15-H1', however, a large NOE is seen between SJG-H11 and A4-H1'(**Figure 94**). This indicates that the 3' end of the drug adduct also exhibits 'S' stereochemistry in accordance with previous studies on inter-strand adducts.[181,195]

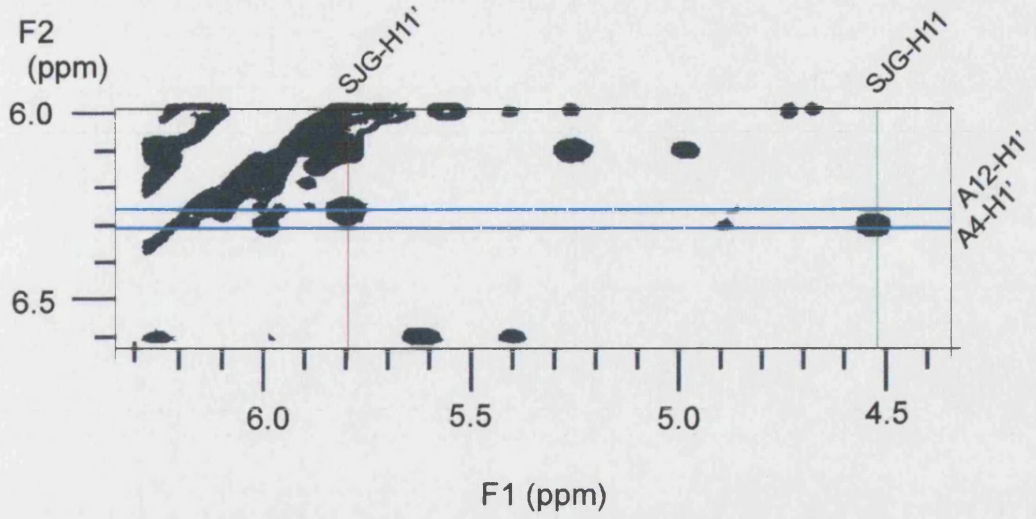
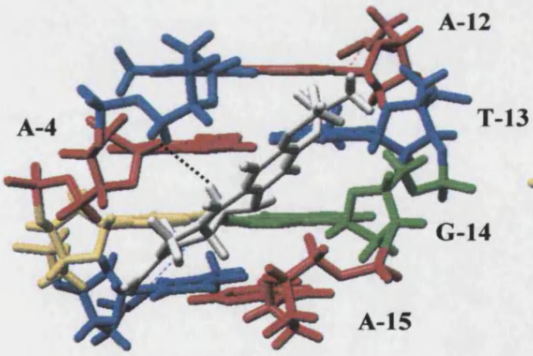
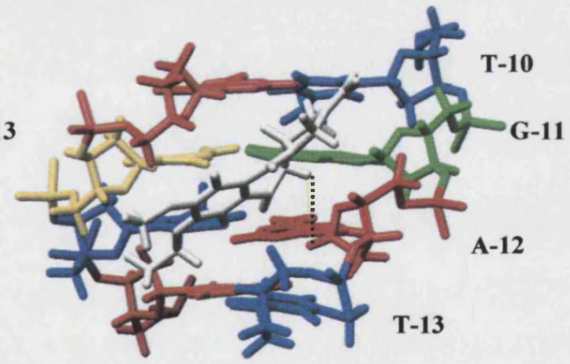


Figure 94: 600 MHz NOESY spectrum of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, 200 ms mixing time. Expansion showing NOE connectivities between SJG-H11/SJG-H11' and A12-H1'/A4-H1' as confirmation of stereochemistry at the SJG-C11 and SJG-C11' centres



'S' stereochemistry at SJG-C11. The SJG-H11- A4-H1' connection is shown by a dotted line



'S' stereochemistry at SJG-C11. The SJG-H11- A12-H1' connection is shown by a dotted line

Figure 95: Representation of part of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, showing NOE connectivities confirming 'S' stereochemistry at both ends. Model produced in UCSF Chimera[193]

4.5: Nucleic Acid Proton Chemical Shifts

As discussed previously, 2D NOESY and COSY experiments were performed on the 5'*d*(GTGATGAG).(CTCATCAC)-SJG-136 adduct in order to determine the proton chemical shifts and intensity of cross-peaks for use in molecular modeling procedures. **Figure 92** shows an expanded NOESY contour plot of the H6/H8 to H1' region of the adduct spectrum. As discussed previously, two complete proton walks of approximately equal intensity can be seen in this region confirming the retention of right-handed B-form conformation of this non-self-complementary adduct.

4.5.1: Relative chemical shifts of the DNA H2' and H2'' protons

In general the H2' protons resonate up-field of the H2'' protons within the same nucleotide sequence.[181] On assignment of the 5'*d*(CICGATCICG)$_2$-SJG-136 adduct in **Chapter 3** it was noted that a reversal in the general pattern for H1'-H2'/H2'' resonances had been observed for the I8 nucleotide, with the more intense H1'-H2' resonance found down-field of the H1'-H2''. This was attributed to a conformational change within the internal nucleotide, with an additional perturbation of the structure of the nucleotide facing the covalently modified guanine. In the case of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct a similar reversal of the general pattern is seen in many of the nucleotides in the vicinity of the drug. This is, perhaps to be expected with the accommodation of SJG-136 in the minor groove as an intra-strand adduct, and perturbation of the surrounding nucleotides provides additional evidence in support of this.

4.5.2: Chemical shift changes of the 5'd(CTCATCAC).(GTGATGAG)-SJG-136 adduct relative to the duplex

Chemical shifts for the oligonucleotide proton resonance signals of the SJG-136 adduct relative to the duplex are shown in **Tables 16 - 18**. Numerous large shifts in resonance can be seen, as would be expected on accommodation of the ligand in the minor groove in this way. The two central adenine H2 resonances (A4-H2 and A12-H2) of the duplex

both experience large down-field shifts of 0.28 ppm & 0.42 ppm respectively, providing further evidence of some central perturbation as a result of formation of an intra-strand adduct. On examination of the H1' and H2' chemical shift data it can be seen that there is a slight majority of large shift changes from duplex to adduct found in nucleotides situated on the covalently modified DNA strand. This suggests that the drug is associating more closely with the modified strand, and as such supports the formation of the intra-strand adduct.

4.6: Intermolecular Drug-DNA Contacts

The proton resonances for SJG-136 have been identified and assigned using analysis of the 2D NOESY and COSY data as discussed in **Chapter 2**. The complete 2D spectra showing assignment of drug proton peaks can be seen in **Figure 93.** It is possible to trace two complete drug proton walks across the entire 2D NOESY spectrum and numerous contacts have been found between the drug protons and the C3, A4, T5, C6 and A7 nucleotides on the non-covalently modified strand, and the G11, A12, T13, G14 and A15 nucleotides on the covalently modified strand. These are shown in **Tables 19 & 20,** as well as pictorially in **Figure 96.** The presence of the two complete drug proton walks (corresponding to the two ends of the drug) as well as the network of NOE contacts from the drug to specific DNA protons once again locates the drug unequivocally within the minor groove and confirms the presence of the intra-strand cross-linked adduct.

4.6.1: Distinction between the non-self-complementary ends of the 5'd(CTCATCAC).(GTGATGAG)-SJG-136 adduct

The two different ends of the SJG-136 molecule were identified by use of a strong NOE contact between the SJG-H9 proton with the respective A-H2 protons of the central adenine nucleotides (A12-H2 and A4-H2) of the DNA duplex. These appear in the H6/H8 – H1' region of the spectrum, and can be seen in **Figure 97.** This assignment is confirmed by the presence of further NOE contacts between the SJG-H9/H9' protons at the respective ends of the drug and the surrounding DNA protons. Examples are SJG-H9 – T5-H1' and SJG-H9' – A12-H1', C6-H1' and T13-H1'. These are shown in **Figures 98 & 99.**

DNA Protons	H1b'	H1a'	H2a'	H2b'	H3a'	H3b'	H6'	H9'	H11'	H11a'	H12a'	H12b'	H13a/b	H14'
A4H2												VW		
C6H1'								M		M/S	S/M	S/M		
C8H1'	M		W											
A7H2	M	S	W	W					W	M				
A7H1'	M	M			W	W				S				
A7H3'														M
A7H4'							M		VW					M/S
G11H1'			S	S										
G11H4'			M	M										
A12H2								S			W/M	W/M	M	
A12H1'								M/W						
A12H4'	M	M	W	M										
T13H1'								M/W			M	M		
T13H4'							W	VW					W	
G14H4'													M/S	

Table 19: Drug-DNA connectivities for the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, G-5' end

DNA Protons	H1a	H1b	H2a	H2b	H3a	H3b	H6	H9	H11	H11a	H12a	H12b	H13a/b	H14
C3H1'			W											
A4H1'						M		W	M/S					
A4H2								S			M/S	M/S		
A4H4'			M	M										
T5H1'								M			M	M		
T5H4'							S							
A12H2											W/M	W/M	M	
T13H1'								W			M	M		
T13H4'													W	
A15H1'	S	S			M									
A15H2	W	M												
A15H4'										M				
G16H4'	VW	VW												

Table 20: Drug-DNA connectivities for the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct, G-3' end

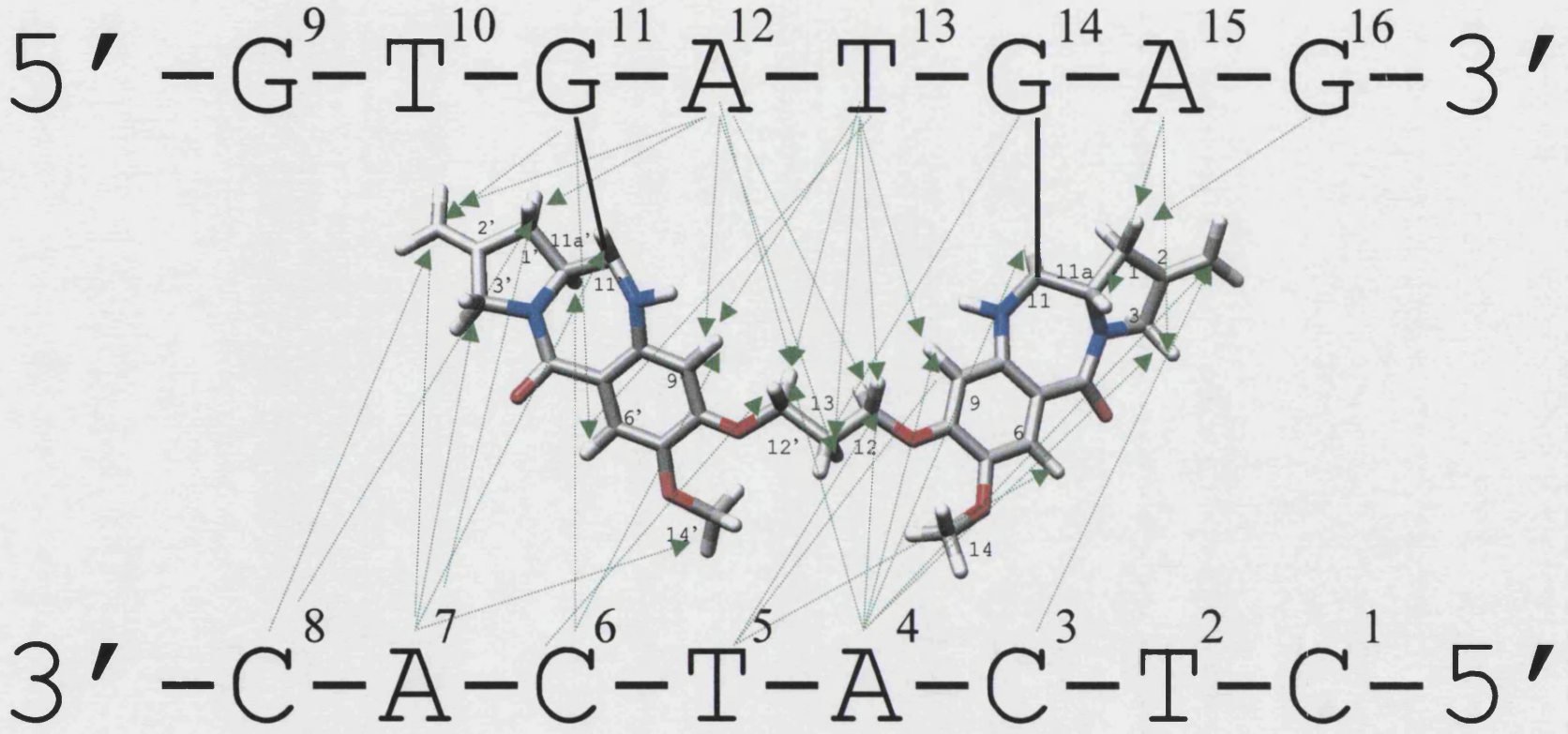


Figure 96: Representation of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct showing NOE connectivities between the drug and the DNA backbone

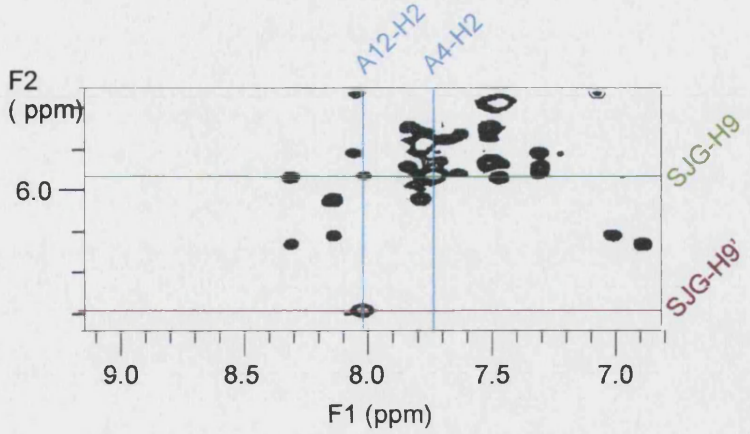


Figure 97: 600 MHz 2D NOESY spectrum of 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136, 200 ms mixing time, expansion showing NOE connectivities between SJG-H9/H9' and A-H2 protons

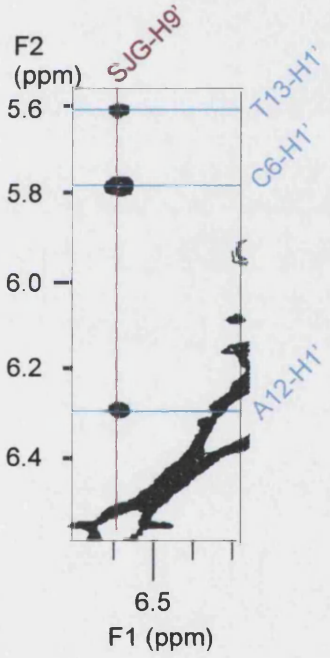


Figure 98: 600 MHz 2D NOESY spectrum of 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136, 200 ms mixing time, expansion showing NOE connectivities between SJG-H9' and DNA H1' protons

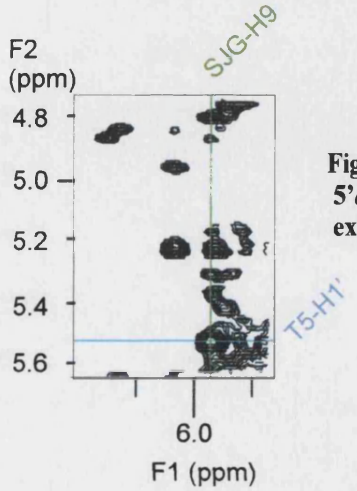


Figure 99: 600 MHz 2D NOESY spectrum of 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136, 200 ms mixing time, expansion showing NOE connectivities between SJG-H9 and T5-H1'

4.6.2: Comparison of the chemical shifts of drug protons in the 5'd(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct relative to the 5'd(CICGATCICG)$_2$-SJG-136 inter-strand adduct

A comparison of the drug chemical shifts for the inter-strand adduct discussed in **Chapter 3** and the intra-strand adduct is shown in **Tables 21 & 22.** It can be seen that at the G-5' end of the duplex the associated SJG-136 protons show a greater similarity in terms of chemical shift. With the exception of the SJG-H1 and SJG-H2 protons, there are no changes greater than 0.12 ppm. This is perhaps a result of a reaction mechanism as discussed in **Chapter 1,** with the alkylation proceeding in two stages.[67] When the chemical shift values of the drug protons located at the 3' end of the modified strand are compared with the inter-strand adduct however large discrepancies are encountered between the two shifts. The SJG-H11 proton resonance in particular has moved by more than 1 ppm, suggesting that the second step of the reaction has resulted in localised perturbation on accommodation of an intra-strand cross-link.

4.6.3: Distinction between 'a' and 'b' protons

Figures 100 & 101 show pictorial representations of the adduct as a stereo-view, identifying protons 'a' and 'b' at positions H1, H2, H3, H12, and H13. Distinction between 'a' and 'b' protons was achieved using the NOE connectivities between the protons and their surrounding DNA protons.

The SJG-H1a/b and H1a/b' protons were assigned on the basis of their respective proximity to the nearby adenine H2 (A7-H2 and A15-H2), as well as through the size of NOE cross-peaks to other drug resonances. When basic molecular models are examined it can be seen that the SJG-H1a proton is closer in space to the A-H2 than the SJG-H1b proton is, and the two were therefore assigned using the size of the resulting NOE cross-peak. In addition SJG-H1a gives rise to more intense cross-peaks than SJG-H1b to SJG-H11 and SJG-H2a/b, reflecting the fact that the SJG-H1a proton is closer in space to these protons than the SJG-H1b proton is.

	H1a	H1b	H2a	H2b	H3a	H3b	H6	H9	H11	H11a	H12a	H12b	H13a	H13b	H14
Inter-strand	2.55	3.21	5.84	5.84	4.02/4.06	4.02/4.06	7.05	6.53	5.64	3.92	4.24	4.24	2.28	2.28	3.83
Intra-strand	2.95	2.58	5.32	5.17	4.32	4.11	6.95	5.97	4.52	4.74	4.25	3.76	2.07	2.30	3.80
Diff (+/-)	*0.40*	*0.63*	*0.52*	*0.67*	-	-	*0.10*	*0.56*	*1.12*	*0.82*	*0.01*	*0.48*	*0.21*	*0.02*	*0.03*

Table 21: Comparison of the chemical shifts of drug protons in the 5'*d*(CICGATCICG)$_2$-SJG-136 inter-strand adduct and the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct – G-3' end. Differences greater than 0.25 ppm are underlined

	H1a'	H1b'	H2a'	H2b'	H3a'	H3b'	H6'	H9'	H11'	H11a'	H12a'	H12b'	H13a	H13b	H14'
Inter-strand	2.55	3.21	5.84	5.84	4.02/4.06	4.02/4.06	7.05	6.53	5.64	3.92	4.24	4.24	2.28	2.28	3.83
Intra-strand	3.22	2.64	5.24	5.22	4.09/4.14	4.09/4.14	7.06	6.58	5.76	3.91	4.20	4.30	2.07	2.30	3.78
Diff (+/-)	*0.67*	*0.57*	*0.60*	*0.62*	-	-	*0.01*	*0.05*	*0.12*	*0.01*	*0.04*	*0.06*	*0.21*	*0.02*	*0.05*

Table 22: Comparison of the chemical shifts of drug protons in the 5'*d*(CICGATCICG)$_2$-SJG-136 inter-strand adduct and the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct – G-5' end. Differences greater than 0.25 ppm are underlined

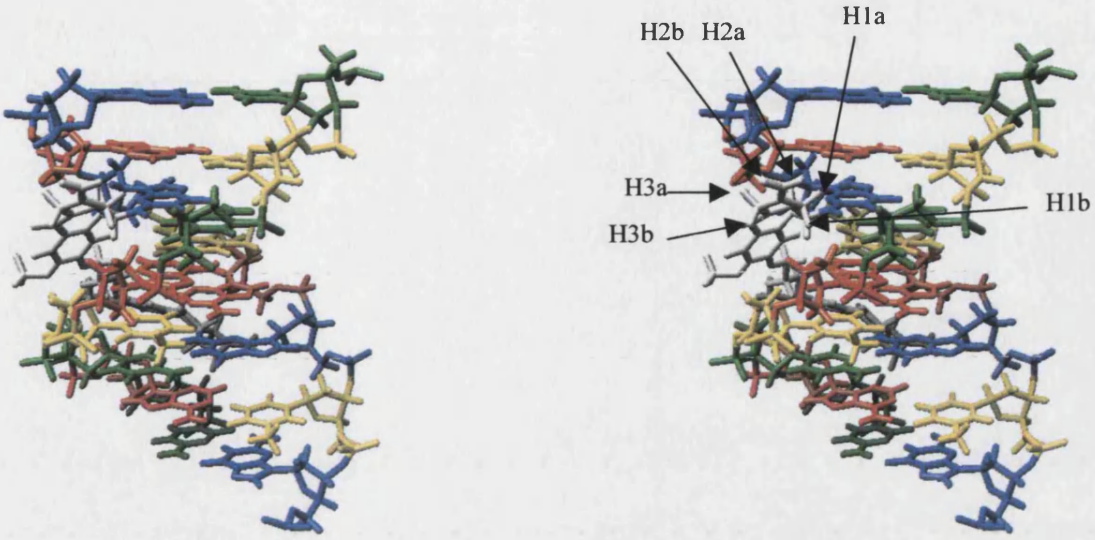


Figure 100: Cross-eye stereo view of an SJG-136-drug adduct, showing numbering of SJG-H1a/b, SJG-H2a/b and SJG-H3a/b protons. Model generated using USCF Chimera[193]

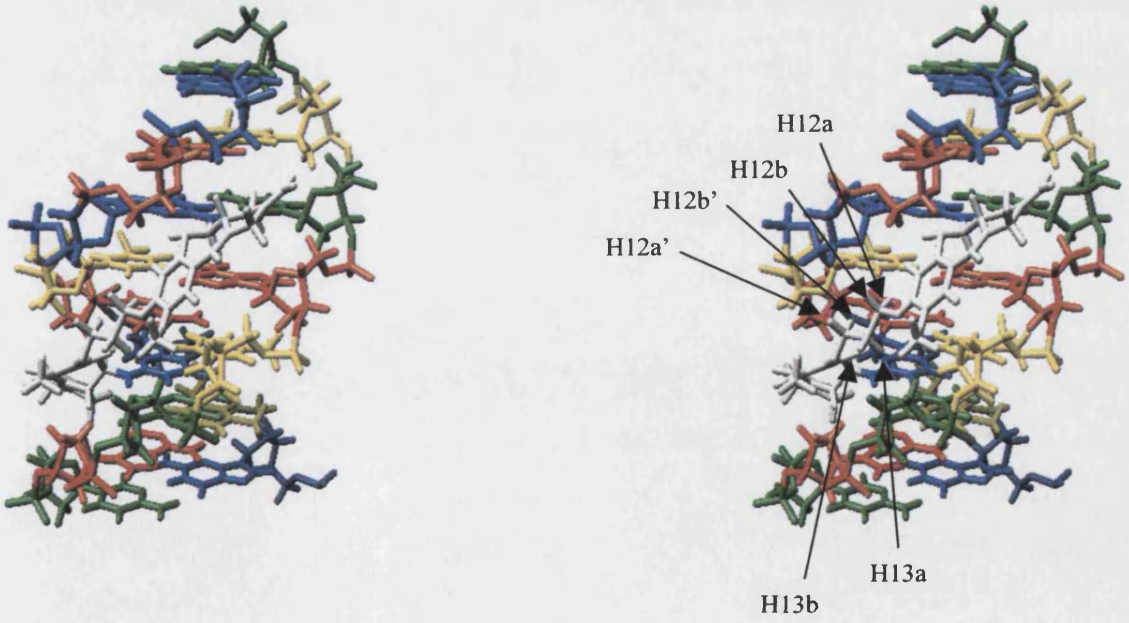


Figure 101: Cross-eye stereo view of an SJG-136-drug adduct, showing numbering of SJG-H12a/b, SJG-H12a/b' and SJG-H13a/b protons. Model generated using USCF Chimera[193]

The SJG-H2a/b protons were similarly assigned based on their proximities to surrounding protons. Unfortunately in this case connectivities to the DNA backbone were too overlaid to be of use, so the SJG-H2a/H2a' and SJG-H2b/H2b' protons were distinguished by their NOE contacts to other drug resonances. When models of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct are examined it can be seen that the SJG-H2a/H2a' proton is closer in space to the SJG-H1/H1' protons, whereas the SJG-H2b/H2b' proton is closer to the SJG-H3/H3' protons. When these NOE connectivities are examined it is observed that the cross-peak intensities vary significantly, and so the SJG-H2a/b/H2a/b' protons were assigned on this basis. This was supported for the SJG-H2a/b by the NOE cross-peaks to the SJG-H11a proton, in which the SJG-H2a proton displayed a slightly more intense resonance.

The SJG-H3a/b protons were distinguished as a result of small differences in spatial proximity to the SJG protons H6 and H2a/b. When distances are calculated from molecular models[193] the SJG-H3a proton is slightly closer to the SJG-H2 protons than the SJG-H3b proton is. In addition the SJG-H3b proton is marginally closer (~ 0.5Å according to models produced within UCSF Chimera[193]) in space to the SJG-H6 proton. When the 2D NOESY NMR spectrum is examined SJG-H3a shows a more intense cross-peak to SJG-H2a/b than SJG-H3b does. On examination of the SJG-H6 connectivities it is observed that there exists a weak NOE to SJG-H3b, while there is no connectivity with SJG-H3a. On this basis a distinction was made between SJG-H3a and SJG-H3b. At the G-5' end of the molecule the SJG-H3a/b' protons were unfortunately too close together and too overlaid to allow a distinction to be made between them. Consequently they are reported as SJG-H3a/b'.

The SJG-H12a/b and H12a/b' protons were once again difficult to distinguish due to their resonance in the overcrowded H4' region of the spectrum. However, they were tentatively assigned based on connectivities to SJG-H9' and adenine H2 protons A12-H2 and A4-H2. The SJG-H12a' proton shows a more intense cross-peak to SJG-H9' than SJG-H12b' does and so a distinction can be made between the two. At the 3' end of the molecule the NOE connectivities of SJG-H12a/b to A4-H2 and A12-H2 show significant differences, with a more intense cross-peak seen due to connectivity with SJG-H12a in both cases.

Finally the SJG-H13a/b protons were distinguished due to their NOE connectivity with SJG-H9'. The distance between SJG-H13b and SJG-H9' is greater than for SJG-H13a, and on examination of the spectrum it can be seen that the cross-peak between SJG-H9' and SJG-H13a is more intense.

4.6.4: Chemical Shifts of T13-H4' and T5-H4'

An interesting feature of the intra-strand 2D NOESY spectrum is the unusually low shift of T13-H4' and T5-H4'. These nucleotides are situated at the centre of the duplex between the two sites of covalent modification. The increased shielding of the 4' protons in these bases is probably due to the proximity of the aromatic PDB rings, and is indicative of some distortion around the SJG-C11 centre. As a two-step reaction has been predicted for alkylation of DNA by PBD ligands[67] this is probably the second centre to react, and some distortion would not be wholly unexpected in the case of accommodation of an intra-strand linkage. In support, there exist no analogous unusual shifts either in the inter-strand adduct reported in **Chapter 3,** or in the inter-strand DSB-120 adduct reported by Mountzouris *et al.*[181]

4.7: The refined molecular model of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct

The 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct was modelled using the **SYBYL**[197] software suite as discussed in **Chapter 3**. Once again drug-DNA adducts were first modelled using a **HYPERCHEM**[198] package. Both DNA and drug were drawn and docked within the package and subjected to steepest descent energy optimisation and secondly to molecular dynamics *in vacuo* at 300K. SJG-136 was then imported from **HYPERCHEM** into the **SYBYL** suite where it was docked with correct stereochemistry onto a DNA duplex produced within the **SYBYL** package. The model was then subjected to procedures identical to those discussed in **Chapter 3**. Complete restraints tables can be seen in **Appendix II.**

Figures 102 - 104 show the *in aquo* computer model of the intra-strand adduct. SJG-136 is depicted in white, with the two DNA strands in magenta and green. As with the inter-strand adduct it is clear that Watson-Crick base pairing has been maintained. This is in agreement with the 2D NMR spectra, which as discussed, suggested that the β-helical structure of the DNA backbone had been maintained.

4.8: Conclusions

In summary, a 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct has been successfully produced and fully assigned. Covalent linkage sites have been confirmed to be the *exo*cyclic NH_2 groups of DNA-G11 and DNA-G14, making this the first positive identification of an **intra-strand** cross-linked DNA-PBD adduct.

The stereochemistry at the SJG-C11 alkylation sites have been confirmed as 'S' at both ends of the duplex, in agreement with molecular mechanics energy calculations, which predict the SJG-C11 'S' stereochemistry to be energetically favoured.[42-44]

Relative chemical shifts of the DNA-H2' and H2'' have been found to show a reversal in the usual pattern for duplex DNA, with the more intense DNA-H1' – DNA-H2' resonance found down-field of the DNA-H1' – DNA-H2'' resonance. This has been noted in previous studies of DNA-PBD adducts, both within this project (**Chapter 3**) and within studies of the DSB-120 adduct[181] and is indicative of a conformational change within the internal nucleotide. As such it was not unexpected on accommodation of an intra-strand cross-linkage.

Intermolecular drug-DNA contacts confirm the presence of the drug located in the minor groove, and following the right-handed contour of the β-helical DNA helix. In addition, NOE connectivities suggest that the SJG-136 molecule associates more closely with the covalently modified strand as expected in an intra-strand cross-linkage.

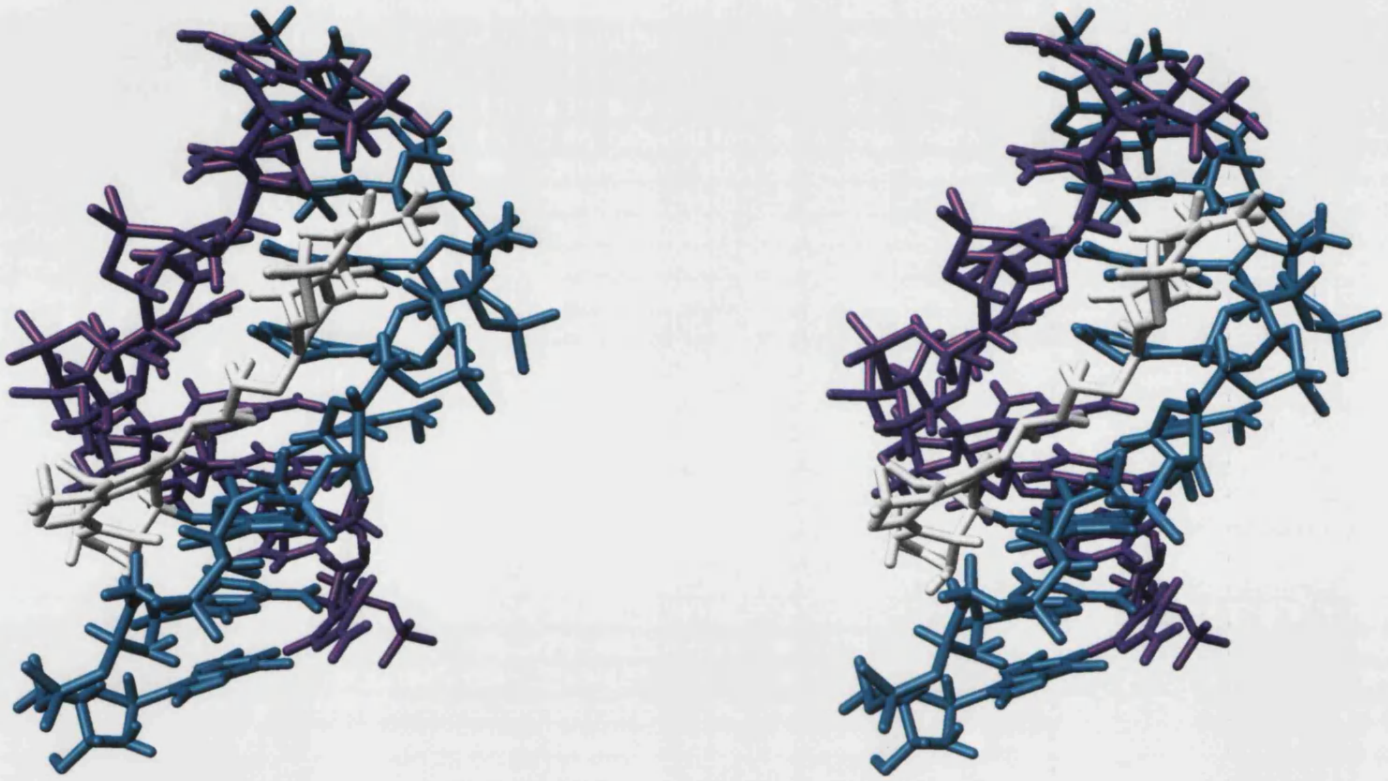


Figure 102: Stereoview of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct. DNA strands are shown in green and purple, SJG-136 is shown in white. Watson-Crick base pairing and the β-helical structure of the DNA backbone have been maintained. Models produced in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

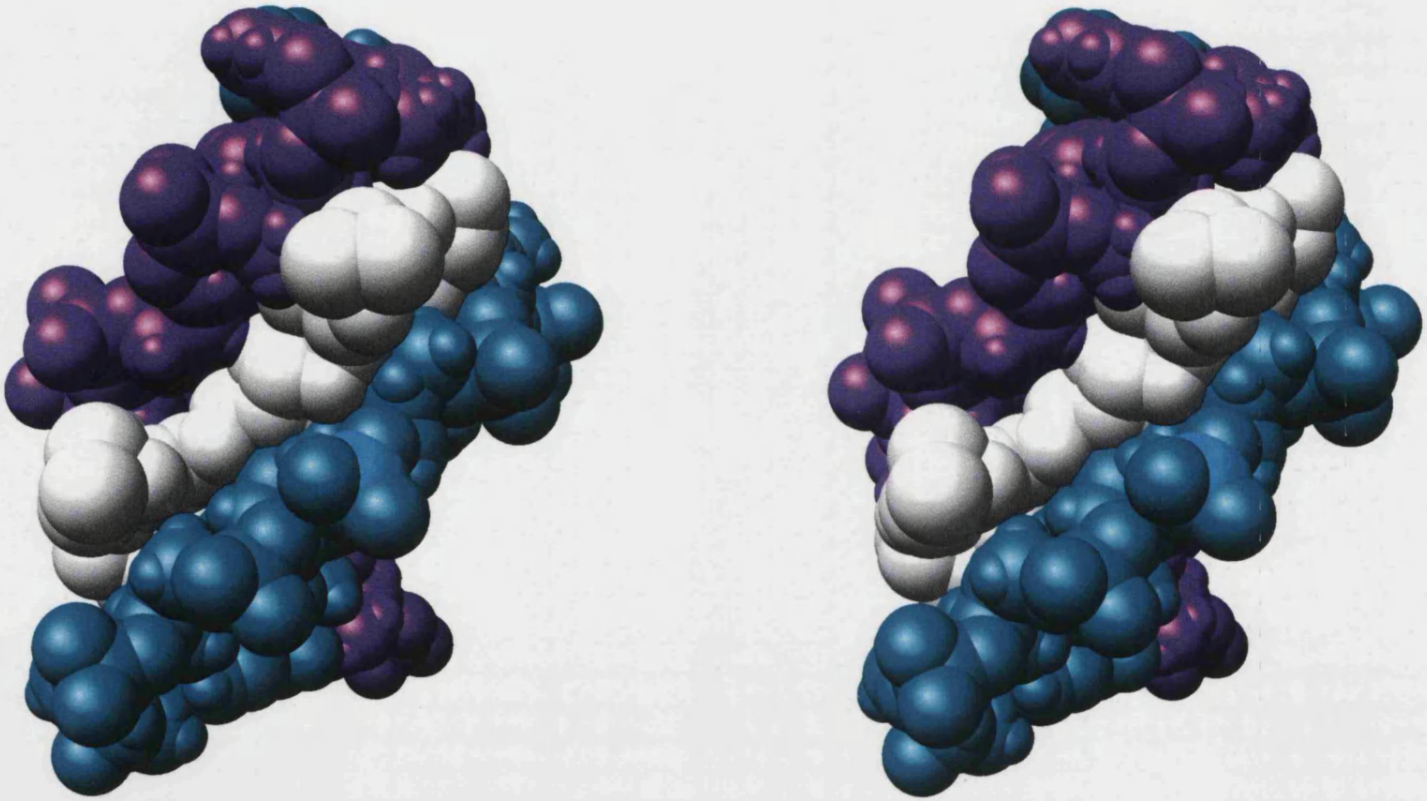


Figure 103: Stereoview of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct. DNA strands are shown in green and purple, SJG-136 is shown in white. The drug is seen to fit snugly into the minor groove with little protrusion beyond the periphery. Models produced in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

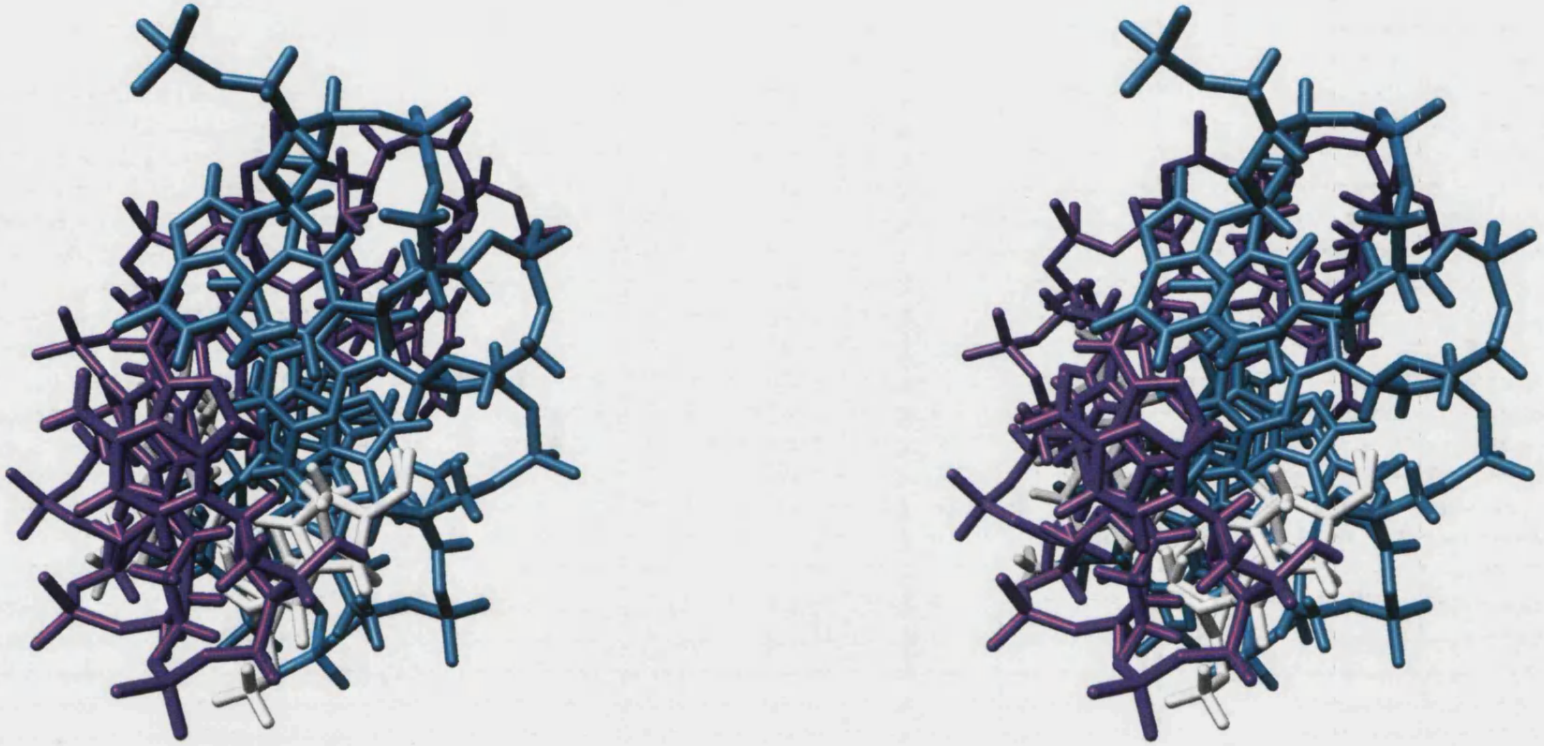


Figure 104: Stereoview of the 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 intra-strand adduct. DNA strands are shown in green and purple, SJG-136 is shown in white. The β-helical structure of the DNA backbone has been maintained. Models produced in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

CHAPTER 5: RESULTS AND DISCUSSION

THE 5'*d*(CGATTAATCG)$_2$-ADOZELESIN ADDUCT

As discussed in **Chapter 1**, adozelesin is a synthetic analogue of (+)-CC-1065, a cyclopropapyrroloindole (CPI) drug containing the DNA alkylating cyclopropa[*c*]pyrrolo[3,2-*e*]-indole-4[5H]one subunit (see **Figure 105**). Adozelesin (**Figure 105**) was synthesised in order to combat delayed lethality problems associated with (+)-CC-1065. It is an overall right-handed molecule, mimicking the pitch of B-form DNA and fitting snugly into the minor groove with a sequence preference of 5'-Pu,Py/Pu,TTA*, where the asterisk denotes the site of covalent modification.

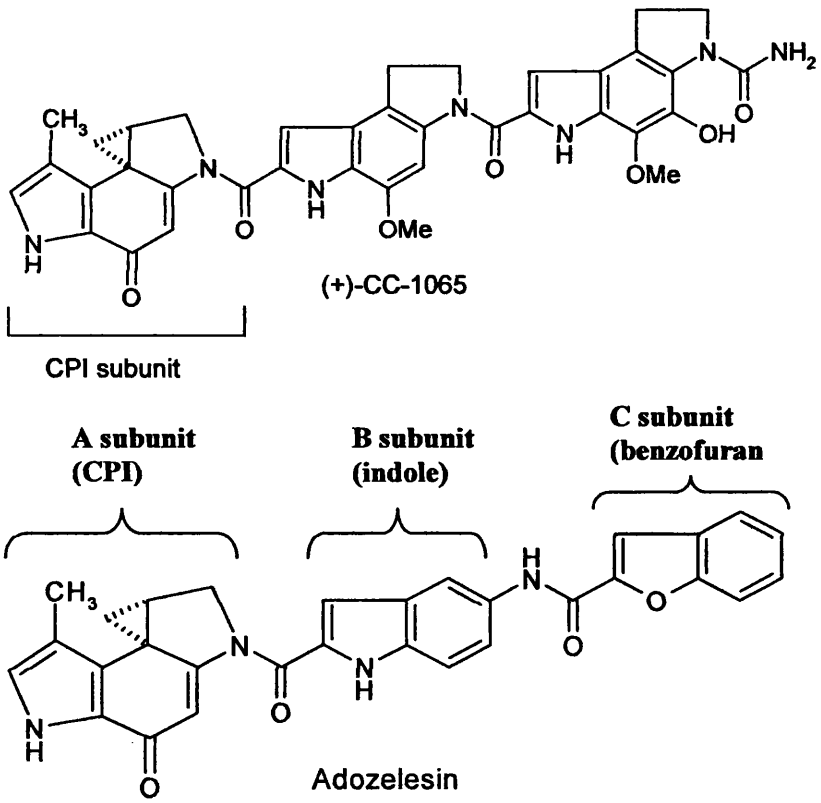


Figure 105: Representation of the (+)-CC-1065 and adozelesin molecules, showing 'A', 'B' and 'C' subunits

Reaction with double stranded DNA is through the N3 position of reactive adenine to form a covalent adduct that overlaps a 5 base pair region in the minor groove.[88-90] The reaction is thought to be a multi-step process utilising non-covalent interaction by hydrophobic and *Van der Waals* forces between the three subunits and the minor groove followed by covalent bonding with the alkylating subunit. The most probable reaction mechanism is shown in **Figure 107.**

In previous work with the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct the adozelesin residues were found to bind to adenine (A12) edge on in the minor groove, with the maintenance of Watson-Crick base pairing. A high degree of hydrogen bonding was found to exist between phenolic protons and the phosphate backbone[104] and a strong hydrogen bond was formed between the amide linker and a thymine (T10) on the modified strand.

In addition to these observations, during molecular modeling experiments[179,180] it was noted that a degree of 'fish tailing' could be observed at the aromatic ring of the 'C' subunit, with the benzofuran tail 'bending' towards the modified strand as depicted in **Figure 106**. As the sequence modelled in these studies was palindromic only half of the adduct was modelled, but it was thought possible that the 'fish-tailing' of the 'C' subunit could be a result of a slight overlap of the aromatic rings of the respective adozelesin molecules in the *bis*-adduct. In this case an increased proximity of the benzofuran tail to the modified strand would be observed as shown in **Figure 106.**

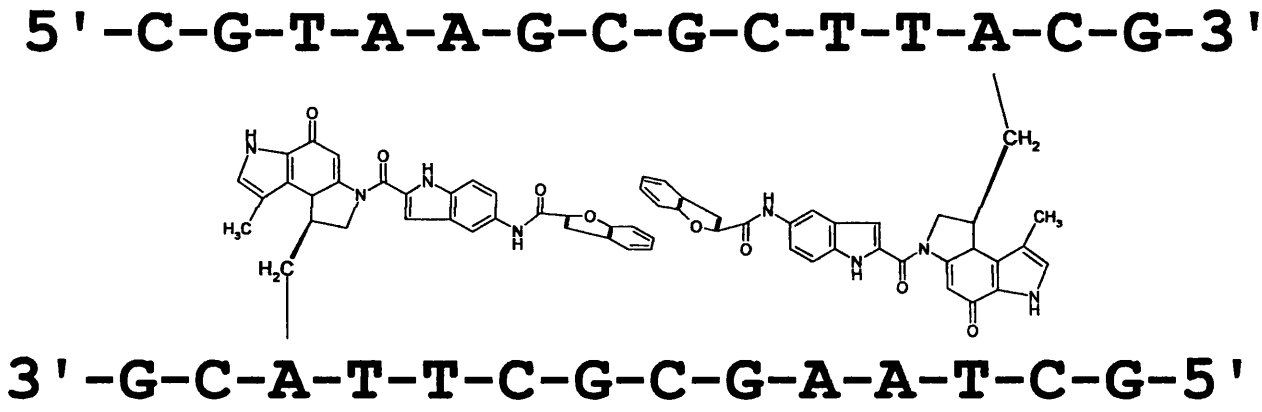


Figure 106: Possible overlap of the 'C' subunit of two adozelesin molecules in the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct

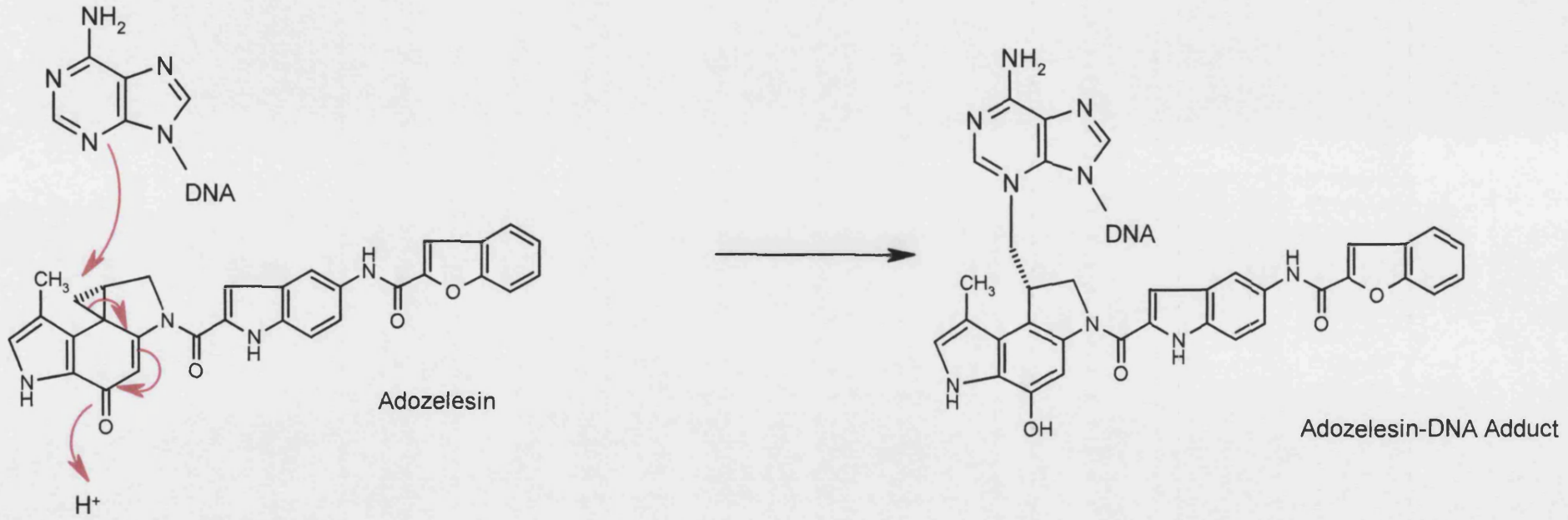


Figure 107: Reaction mechanism of adozelesin with the adenine N3 of duplex DNA

5.1: Design and synthesis of the duplex

The 5'*d*(GCTAATTAGC)$_2$ duplex was designed to investigate the possibility of an overlap of the aromatic rings of the 'B' and 'C' subunits in the DNA-adozelesin adduct. Adozelesin is known to span five base pairs so the new DNA sequence was designed to be four base pairs shorter than that used by Cameron and Thompson,[179,180] in order to attempt to force an overlap between drug molecules in the resulting adduct as shown in **Figure 108.**

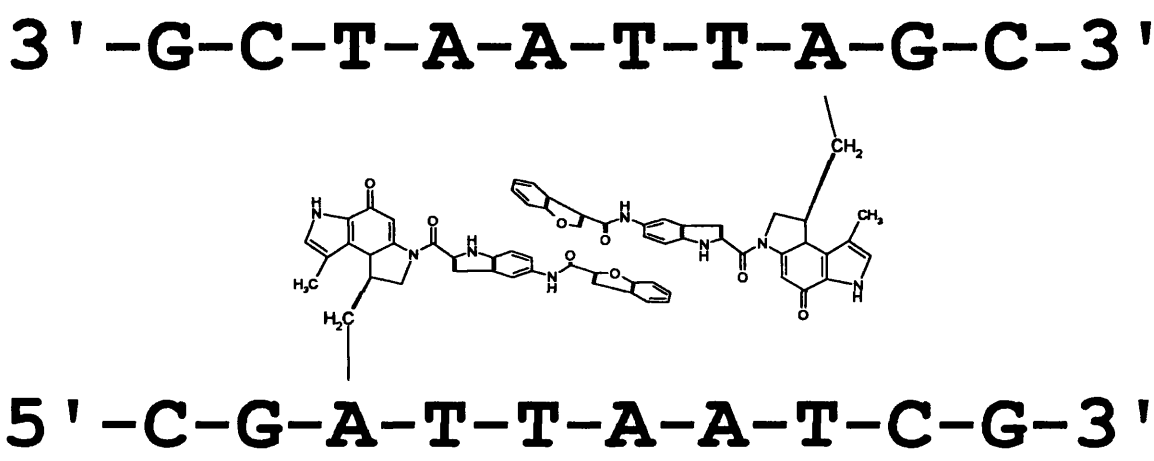


Figure 108: Possible overlap of adozelesin 'B' and 'C' subunits in the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct

The most feasible reaction site for adozelesin with this duplex would be through the N3 position of the adenine residue A3. In order for a *bis*-adduct to form significant overlap would be required between the 'B' and 'C' subunits of the respective adozelesin molecules.

As for previous duplexes the 5'*d*(GCTAATTAGC)$_2$ duplex was synthesised using standard methods as discussed in **Chapter 2.** The duplex was then allowed to react with 4 mg of adozelesin for 48 hours at room temperature. On observation of the 1D ^{1}H NMR spectrum at this point the reaction appeared to be incomplete, and so the reaction

was left to proceed for a further 10 days, at which point it was deduced that the reaction had proceeded to completion. The rate of this reaction was in contrast to that observed by Cameron and Thompson with the longer DNA duplex, in which the reaction was complete after 48 hours.[179] On completion of the drug-DNA reaction 2D NOESY and COSY experiments were performed on the resulting adduct. A diagram showing the numbering system for the duplex for the purposes of NMR assignment is shown in **Figure 109.**

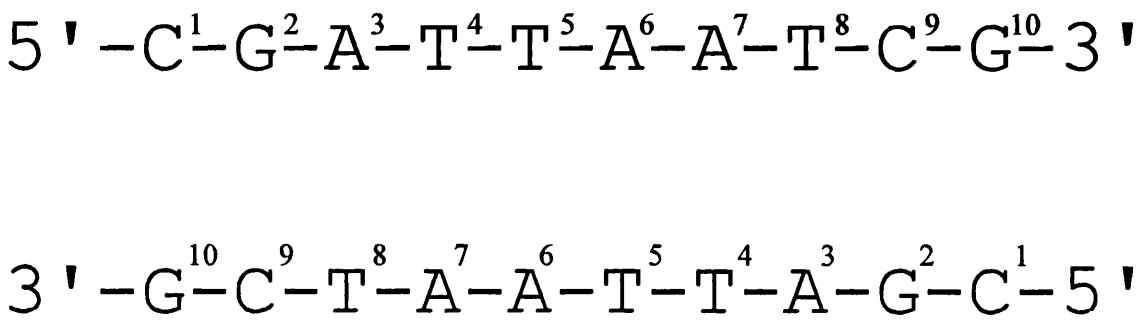


Figure 109: Numbering system for the 5'*d*(CGATTAATCG)$_2$ duplex

5.2: Symmetry of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct – identification of two symmetrical adducts.

The one-dimensional ^{1}H NMR non-exchangeable spectra for both the duplex and the adduct can be seen in **Figures 110 & 111.** The duplex signals are well-resolved, and a peak count in the aromatic region affords approximately 15 peaks – as would be expected for this duplex. This information, along with the three clear thymine methyl peaks that can be observed between 1.00 ppm and 1.60 ppm provides evidence for the purity of this duplex.

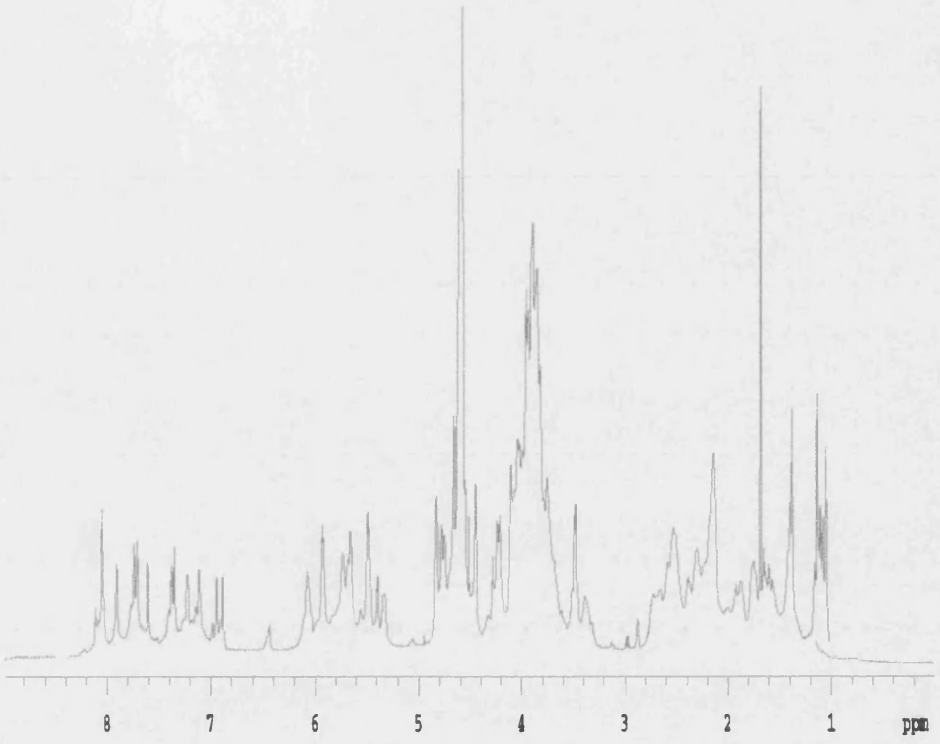


Figure 110: ^{1}H NMR spectrum of the 5'*d*(CGATTAATCG)$_2$ duplex

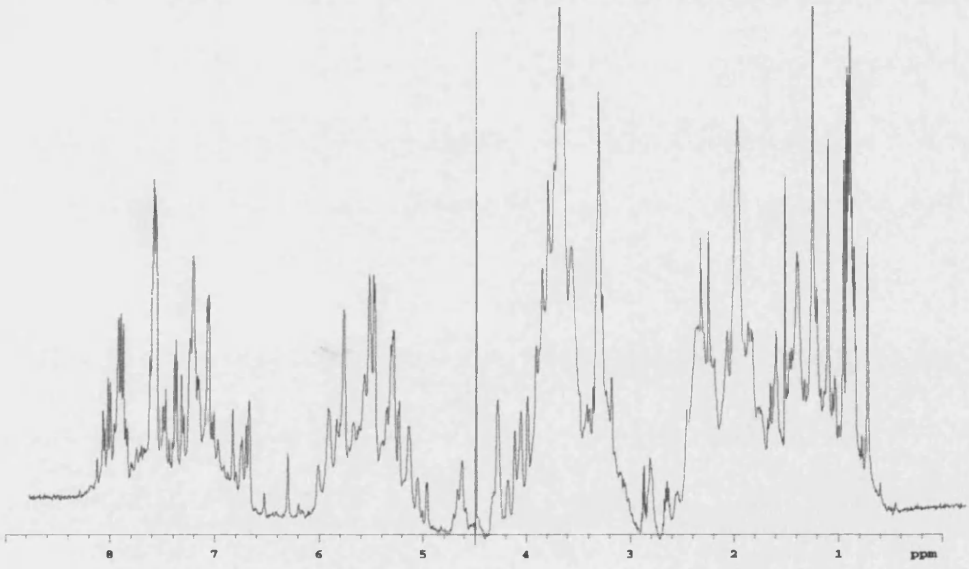


Figure 111: ^{1}H NMR spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct

The adduct spectrum is greatly increased in complexity and due to overlaid signals it is not possible to complete an accurate peak-count in the aromatic region. However on estimating the number of peaks it is apparent that the numbers in the aromatic region as well as the thymine methyl region have more than doubled, and it was deemed possible that more than one species was present in the reaction mixture.

The 5'*d*(CGATTAATCG)$_2$-adozelesin adduct was further investigated using COSY and NOESY NMR techniques. An expansion of the 2D NOESY adduct and duplex spectra, H6/H8-H1' region, are shown in **Figure 112**. On superimposition of the duplex 2D NOESY spectrum on the adduct spectrum it was deduced that the reaction had proceeded to completion as there was no evidence of remaining duplex cross-peaks. For example, the duplex cross-peak at coordinates (F1/F2) 6.96 ppm/5.97 ppm has completely disappeared while a new peak is evident at coordinates 8.15 ppm/6.24 ppm. However, the presence of four strong COSY resonances in the H6/H8-H1' region confirms the presence of two separate species in the reaction mixture. **Figure 113** shows the region completely assigned and it can be seen that two complete proton walks are possible using the sequential assignment method from H1' resonances to H6/H8 resonances. As there is no duplex DNA remaining in the mixture this confirms that two adducts are present, and that both adduct species are self-complementary, retaining B- form character.

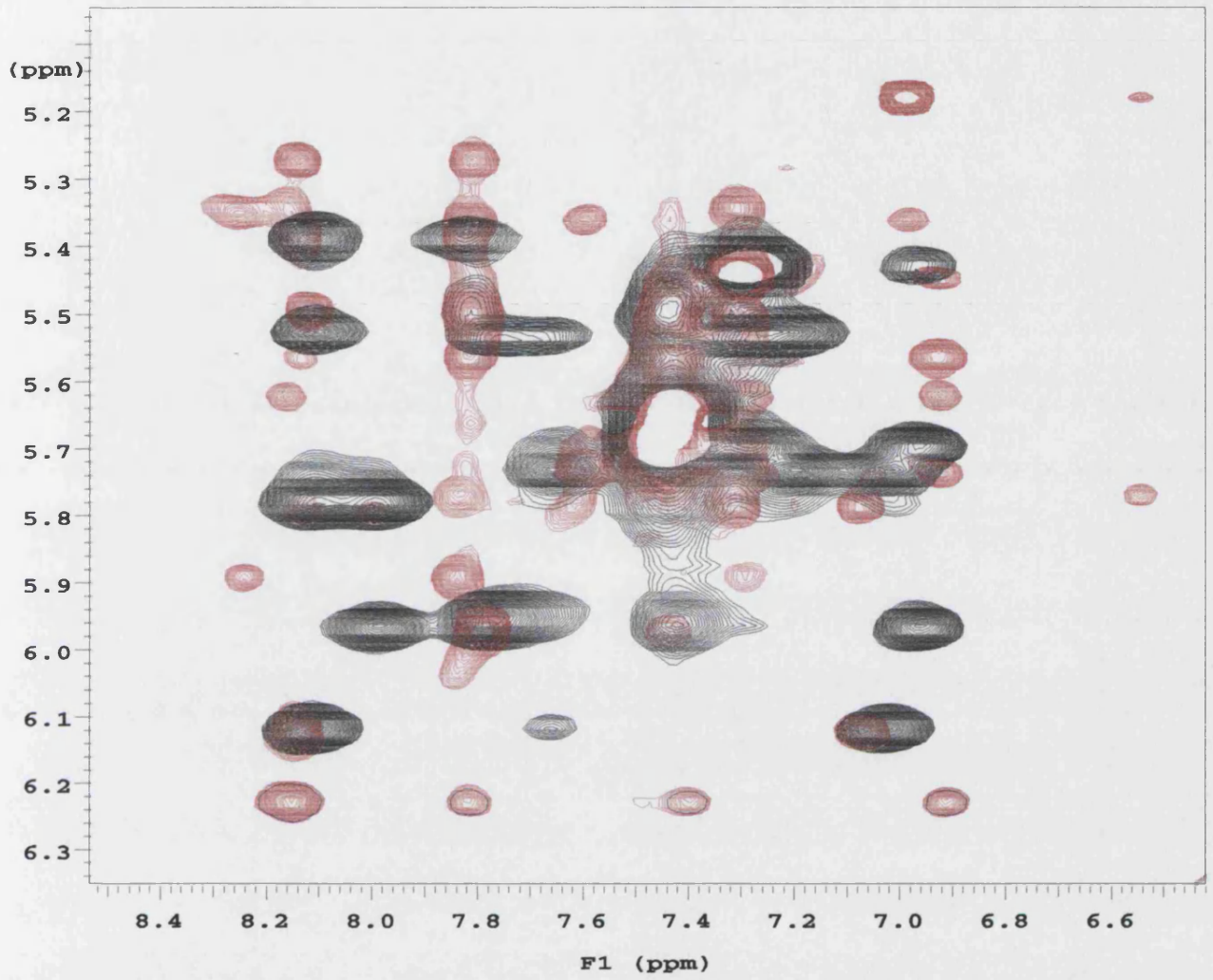


Figure 112: 600 MHz 2D NOESY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct and 5'*d*(CGATTAATCG)$_2$ duplex, H6/H8-H1' region. Mixing time 200 ms. Black=duplex, Red=adduct

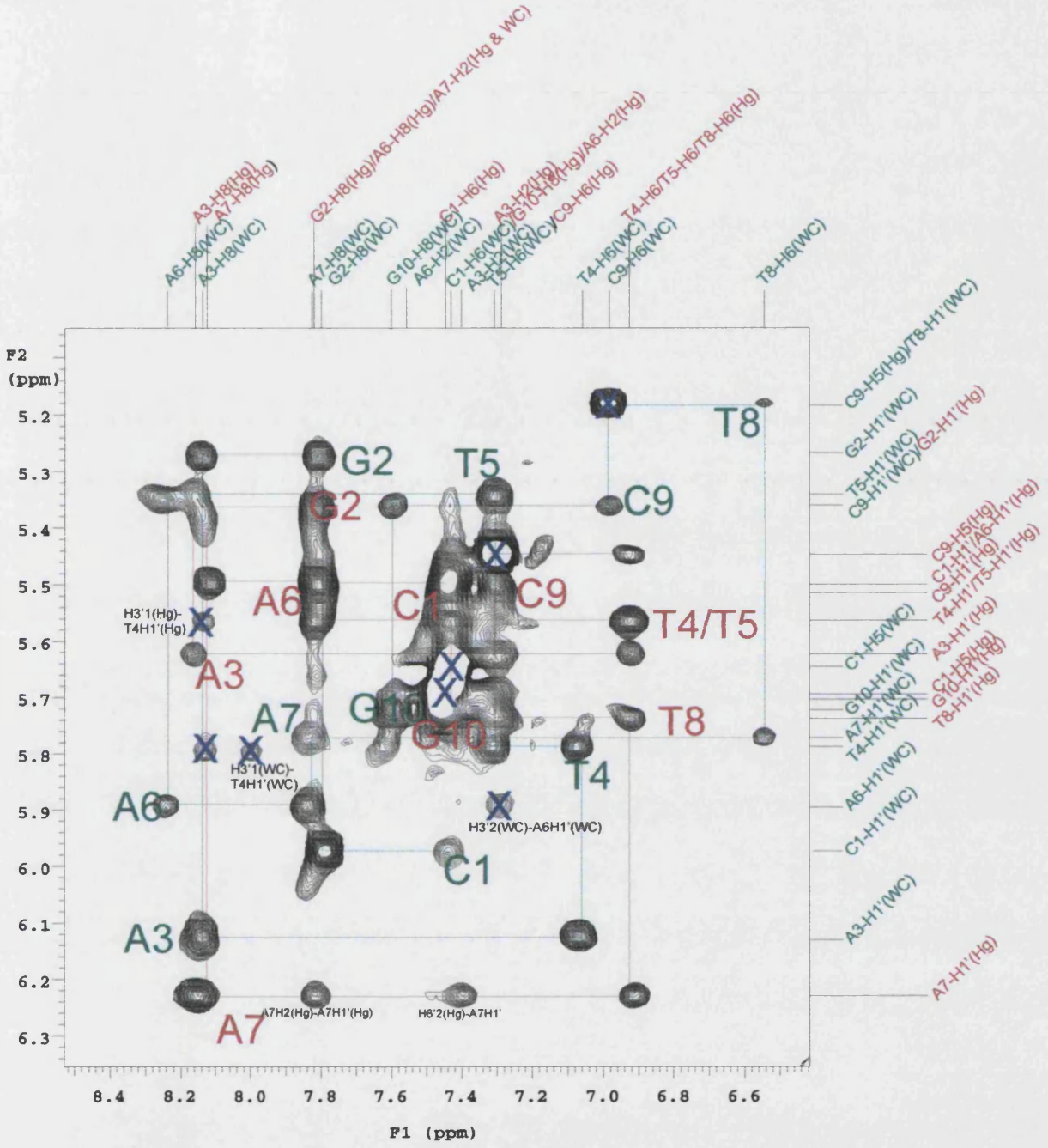


Figure 113: 600 MHz 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct, 200 ms mixing time. Expansion of the H6/H8-H1' region showing two complete walks. Green = Watson-Crick (denoted WC) Red = Hoogsteen (denoted Hg) X = COSY (unlabelled) and peaks due to drug interaction (labelled)

5.3: Identification of the covalent sites in both DNA-adozelesin adducts

As discussed previously, adozelesin is known to react through the N3 position of a reactive adenine base, spanning a distance of five base pairs. As such, on reaction with the 5'*d*(CGATTAATCG)$_2$ duplex the adozelesin molecule has two possible reaction sites on each duplex strand - either at base A3 or base A6 as shown in **Figure 114.**

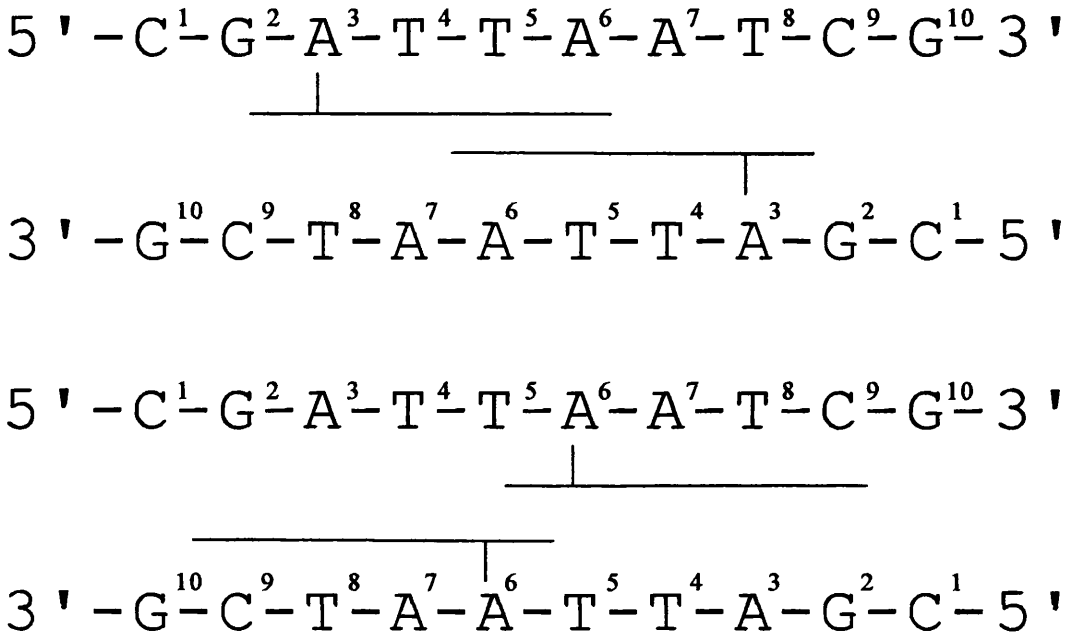


Figure 114: Possible reaction sites for adozelesin on the 5'*d*(CGATTAATCG)$_2$ duplex

Reaction at base A6 is unlikely due to the hydrophobic nature of the adozelesin molecule. In the case of reaction at A6 the benzofuran tail (subunit 'C') of the molecule would be left relatively exposed, as opposed to the snug fit in the minor groove afforded by reaction at A3. In addition large steric clashes would also occur between the 'A' subunits of the respective molecules and as such the A3 nucleotide is the most likely site of covalent modification.

As previously discussed, alkylation of DNA has been observed to proceed through opening of the cyclopropyl ring and formation of a C8 methylene bridge. This is confirmed in this case by the observation of down-field shifts of the adozelesin cyclopropyl ring protons Ado-H8A and Ado-H8B. The shifts of Ado-H8A and Ado-H8B would be expected at 1.47 ppm and 2.02 ppm respectively.[200] In the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct these proton resonances are shifted to the methylene region, at 5.50 ppm in the case of the first adduct and 5.35 ppm for the second. This is also in analogy with the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct in which the Ado-H8A and Ado-H8B proton resonances were found at 5.06 ppm and 5.15 ppm respectively,[104] and confirms that the covalent bond with adenine has formed *via* the C8 methylene bridge.

Further evidence exists for DNA-A3 as the site of covalent modification in the observation of NOE connectivities from the A-H2 proton of the modified adenine into the CPI headunit. NOEs can be found from A3-H2 into Ado-H8A2, and Ado-H1A/B in the case of both adducts. A lack of any cross-peak between A3-H2 and the adozelesin H3 proton is in analogy with previous work on the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct and indicates that the drug is bound in an edge-on orientation in the minor groove.

5.4: Possible adduct conformations

Analysis of the 2D NOESY and COSY spectra confirms the presence of two different DNA-drug adducts. There are four clear COSY cross peaks visible at F1/F2 coordinates 7.44 ppm/5.70 ppm, 7.44 ppm/5.65 ppm, 7.31 ppm/5.45 ppm, and 6.98 ppm/5.18 ppm confirming the presence of four distinct cytosine resonances. In addition two complete proton 'walks' are possible in the H6/H8 – H1' region of the NOESY spectrum (**Figure 113**). Comparison of the duplex NOESY spectrum with that of the adduct showed no evidence for residual duplex in the reaction mixture as shown in **Figure 112,** and two drug walks are present across the whole spectrum (seen in **Figures 115 & 116**). The drug resonances in both adducts display cross-peaks that confirm their close association with DNA duplex.

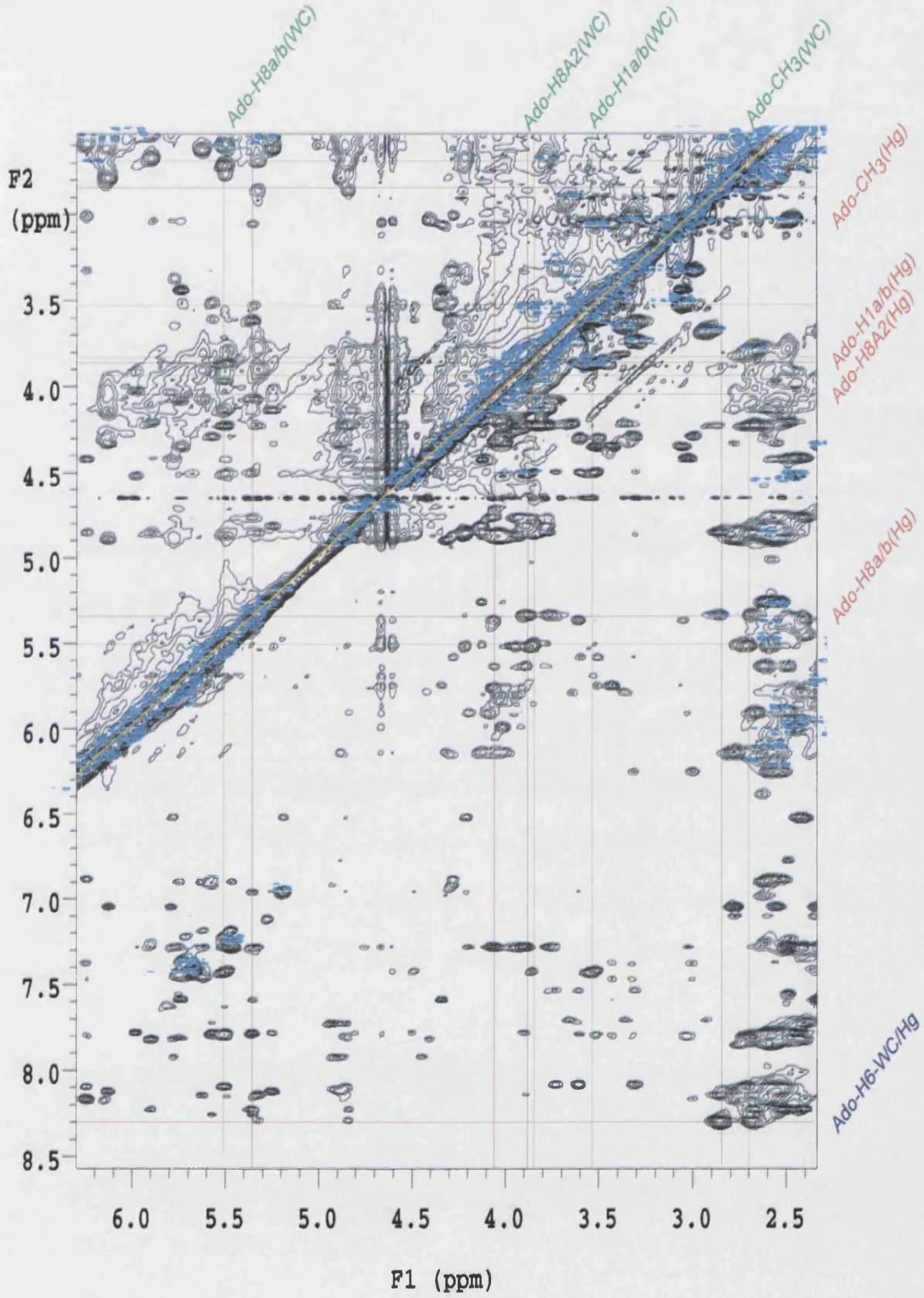


Figure 115: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. NOESY mixing time 200 ms. Black=NOESY, Blue = COSY. Green lines = Watson-Crick assignments, Red lines = Hoogsteen assignments

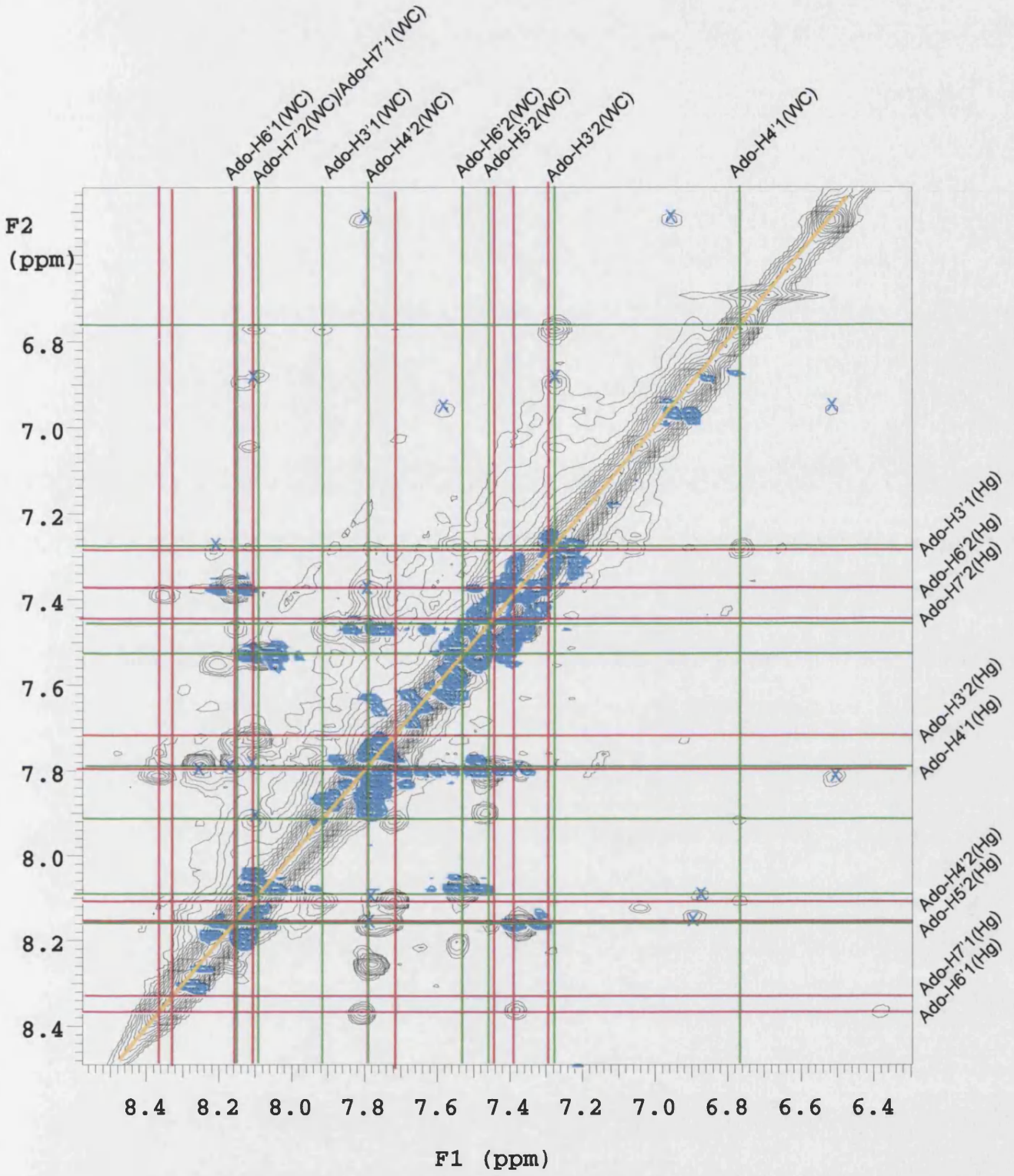


Figure 116: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. Black = NOESY, Blue= COSY. NOESY mixing time 200 ms. Expansion of the H6/H8-H6/H8 region. Green lines = Watson-Crick assignments, Red lines = Hoogsteen assignments

5.4.1: Possible mixed adduct species

Six possibilities exist to explain the presence of two separate proton walks in the 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct. The first is that there is simply unreacted DNA duplex remaining in the reaction mixture. As discussed above this can be ruled out on the basis of a comparison between the 2D NOESY spectrum of the duplex and that of the mixture after reaction with adozelesin. When the two spectra are superimposed it is clear that no duplex peaks remain. **(Figure 112)**

The second possibility is that of alkylation at a different adenine residue – namely A6 or A7. Once again this possibility is not feasible in the reaction of adozelesin with this duplex. The site of covalent modification in the case of both adducts has been confirmed to be at A3. In addition, reaction at A6 or A7 to form a *bis*-adduct would require an overlap of the 'A' and 'B' subunits of the drug. These units are not planar and any overlap would be subject to huge steric clashes and hence highly unfavourable.

A third possibility exists in the formation of a mono-adduct alongside the expected *bis*-adduct, such as that depicted in **Figure 117.**

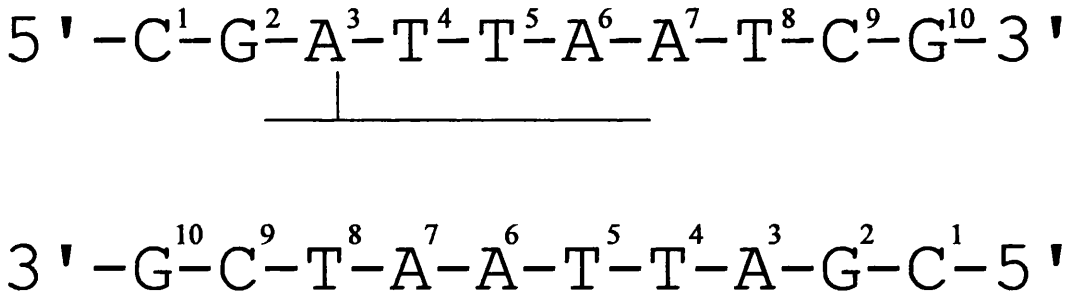
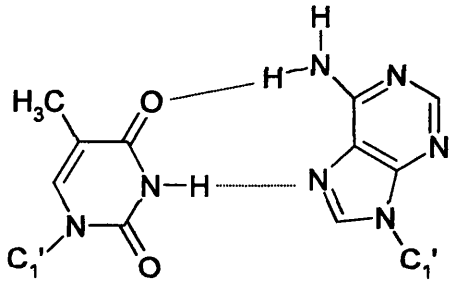


Figure 117: Representation of a 5'*d*(CGATTAATCG)$_2$-adozelesin mono-adduct

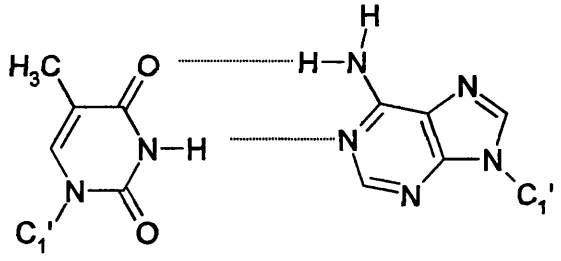
Examination of the 2D NOESY spectrum of the reaction mixture again renders this possibility extremely unlikely. Observation of the two completed walks in the H6/H8-H1' region confirms that both species present are symmetrical. In the case of a mono-adduct the symmetry of the species would be lost as the two DNA strands would no

longer be equivalent. In addition, the unmodified DNA strand would closely resemble duplex DNA, particularly at the terminal ends of the strand. As discussed there is no evidence for duplex resonances in the spectrum so it is unlikely that anything resembling a duplex remains in the reaction mixture. In addition, the time taken for the reaction to complete is vastly greater than that observed in previous studies of CPI-DNA adduct formation.[104]

There remain three possibilities for identification of the two adducts present. Firstly, the formation of a Watson-Crick *bis*-adduct with overlap of the 'C' subunits as predicted. Secondly formation of a base-paired overlapping adduct in which the central AT step displays Hoogsteen base pairing **(See Figure 118)** and finally formation of an overlapped *bis*-adduct with open base pairs.



Hoogsteen A-T base pairing



Watson-Crick A-T base pairing

Figure 118: Traditional Watson-Crick and Hoogsteen base pairing

As discussed in **Chapter 1**, previous work involving reaction of a 5'*d*(CGTAATTACG)$_2$ duplex with the dimeric CPI bizelesin resulted in mixed adduct formation in which approximately 60 % consisted of an adduct in which both adenine

residues of the central AT step were Hoogsteen base paired to thymines (see **Figure 119**), while the remaining 40 % existed in an open conformation.[115] It is suggested that the insertion of bizelesin into the minor groove provides impetus for reorientation of the central adenine residues and stabilises the resulting Hoogsteen base pairs upon completion of cross-linkage.

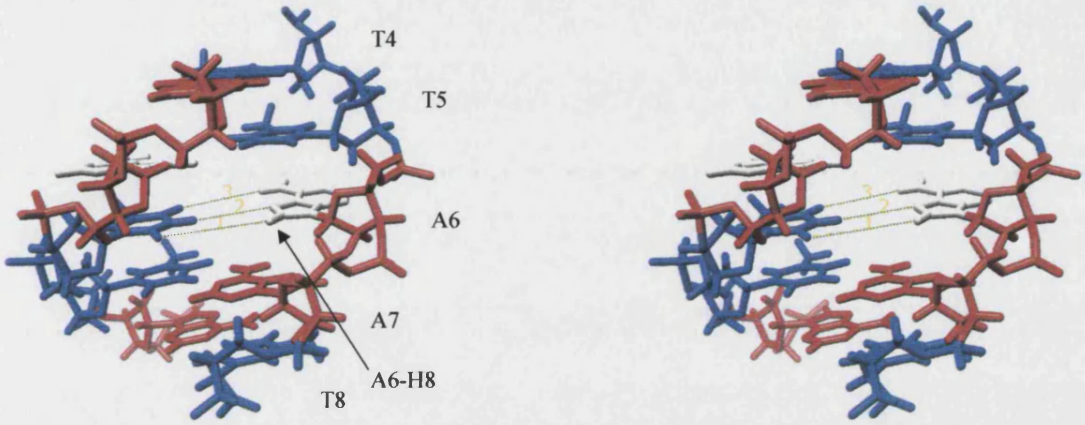
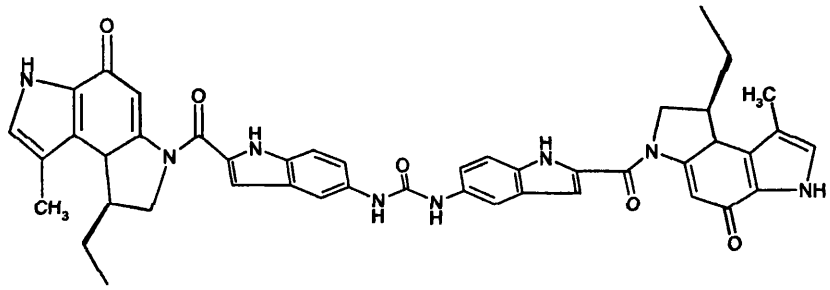
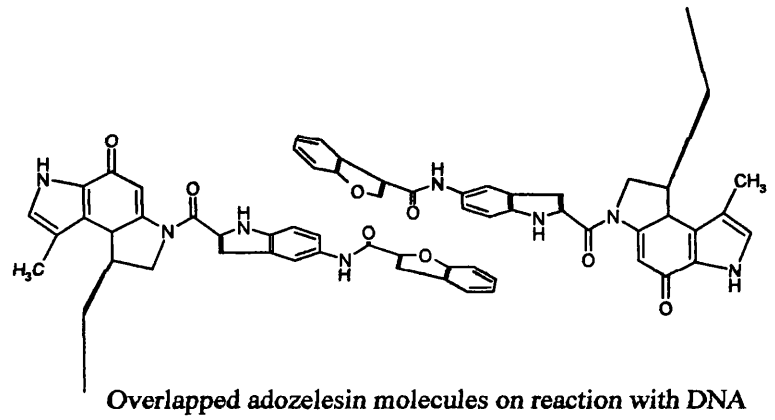


Figure 119: Representation of a DNA duplex showing Hoogsteen base pairs. Hoogsteen bases are shown in white, with hydrogen bonds in green. It can be observed that the A-H8 proton is relocated into the minor groove

In the case of the 5'*d*(CGATTAATCG)$_2$–adozelesin adduct the overlap of the 'C' subunits provides a similar shaped drug to the dimeric bizelesin as shown in **Figure 120**, and it is feasible that the Hoogsteen adduct is 'frozen' out in the same way as observed in the study of bizelesin.



Bizelesin molecule on reaction with DNA



Overlapped adozelesin molecules on reaction with DNA

Figure 120: Representation of bizelesin and overlapped adozelesin after reaction with DNA, *showing similarities in shape*

The Hoogsteen rotation is a stable duplex answer to the steric stress associated with cross-linkage[115] and as the overlapped region of the adozelesin drug moieties unquestionably result in a large increase in steric crowding at this region it is perhaps unsurprising that the duplex DNA should adopt this conformation as a solution. Quinoxaline intercalators have been shown to stabilise partially Hoogsteen base pairs to the 3' side of intercalation[201,202] but the intercalator insertion disrupts base stacking and base-backbone interactions in the Hoogsteen region. Overlapped adozelesin molecules however, like bizelesin, will not intrude into the base stacking of the Watson-Crick to Hoogsteen base pair transition region and therefore is capable of yielding stable duplex

DNA in which the central base pairs show Hoogsteen pairing in a similar way to that observed in studies of bizelesin.

The possibility of open base pairs was investigated using 2D ROESY spectroscopy. As discussed in **Chapter 2,** ROESY experiments are used to determine conformational exchange on an NMR timescale. Exchange within a molecule is observed as peaks of an opposite sign in the NMR spectrum. The ROESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct showed no evidence for conformational exchange and the possibility of open base pairs was ruled out as a result. It was therefore concluded that the mixed adozelesin adduct consisted of two adducts, one maintaining traditional Watson-Crick base pairing throughout, while the other exhibits formation of Hoogsteen base pairs at the central AT step in analogy with previous studies on bizelesin. The two species are denoted adozelesin-Watson-Crick (WC) and adozelesin-Hoogsteen (Hg) in all subsequent discussion.

5.4.2: Orientation of the adozelesin molecules in the Watson-Crick and Hoogsteen adducts

There are two possible overlap orientations for both the Watson-Crick and the Hoogsteen adduct. A representation of these is shown in **Figure 121 A & B.** It can be confirmed by NOESY connectivities from the 'B' and 'C' subunits of the drug molecules to the modified DNA strand that a closer association is formed between the drug and the modified strand in both cases, and so the orientation can be confirmed to be that shown in **Figure 121 B.** Further discussion of this topic can be found in **section 5.6.1.**

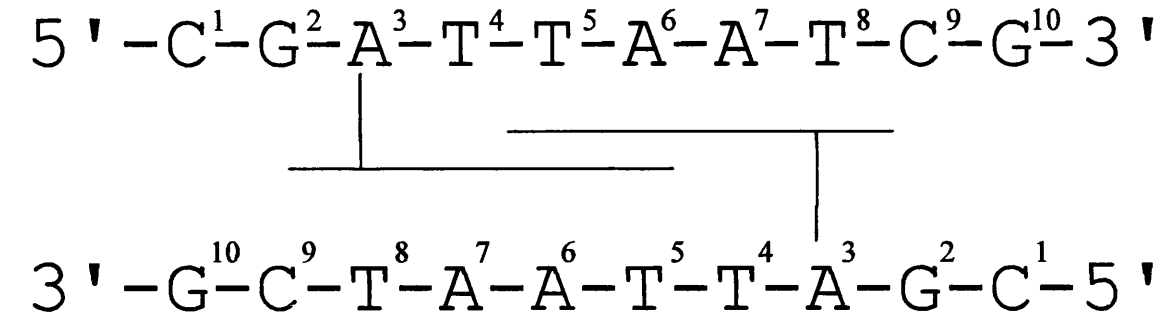


Figure 121 A: Possible orientation for the 5'*d*(CGATTAATCG)$_2$-adozelesin overlapped adduct – each drug molecule associates more closely with the opposite strand

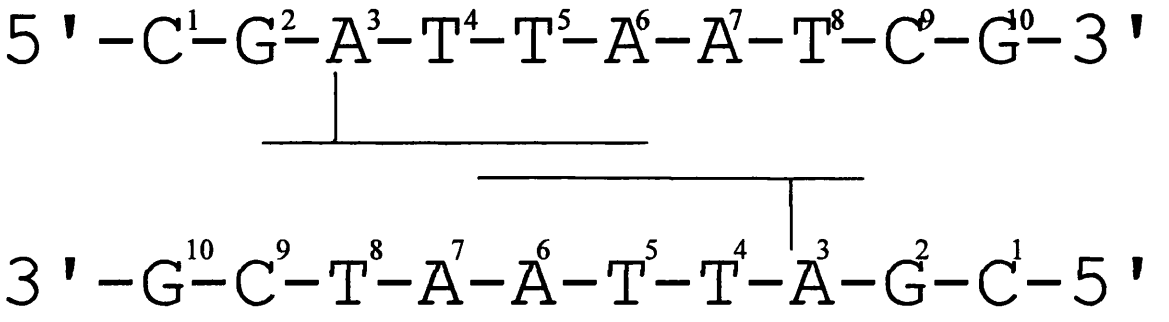


Figure 121 B: Second possibility for orientation of the 5'*d*(CGATTAATCG)$_2$-adozelesin *overlapped adduct – each drug molecule associates more closely with its own DNA strand*

5.5: Nucleic Acid Proton Chemical Shifts – Watson-Crick

The 5'*d*(CGATTAATCG)$_2$-adozelesin adduct was examined by 2D NOESY and COSY NMR and proton resonances were assigned using the assignment protocol discussed in **Chapter 2**. The complete spectra are shown in **Figures 116 & 117** while **Figure 113** shows an expanded NOESY contour plot of the H6/H8 - H1' region of the spectrum. As discussed in **section 5.2** two complete walks of approximately equal intensity are possible in this region, leading to the deduction of the presence of two *bis*-adducts with an overlap of the 'C' subunits of each adozelesin molecule. Both adducts retain self-

complementarity and B-form DNA character, but whilst one adduct maintains Watson-Crick base pairing the second adduct displays a rotation of the adenine residue of the central AT step in each strand – forming a Hoogsteen base pair. This is in analogy with previous studies on adducts of the dimeric CPI drug bizelesin[115]

Proton chemical shifts in the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct relative to duplex DNA can be found in **Table 23.** Down-field shifts of the adozelesin cyclopropyl ring protons Ado-H8A and Ado-H8B (5.50 ppm) relative to unbound adozelesin (1.47 ppm and 2.02 ppm respectively[200]) confirms the site of covalent attachment as C8 of the adozelesin molecule as expected. Large up-field shifts are observed in the T8-H6 nucleotide aromatic and T8-H1' protons, which are situated directly opposite to the modified A3 base. The above-mentioned resonances can be seen in **Figure 122** and it is noted that they are also extremely weak relative to other DNA aromatic and H1' resonances. This suggests a degree of backbone distortion as a result of the overlapped region of the adozelesin molecules. It is possible that the drug molecules are required to 'push back' in order to achieve π stacking between the aromatic rings. This movement of drug molecules in the narrow minor groove causes distortion along the backbone of the DNA duplex, which centres on the T8 nucleotide. This is supported by further down-field shifts of 0.68 ppm at the aromatic proton of the C9 nucleotide (C9-H6), indicating T8 and C9 as the probable centre for any distortion. The presence of complete proton 'walks' indicates that the β-helical structure is maintained, but the weakness of resonances connected to the T8 nucleotide show that the molecular distance is at the limit of detection. Further molecular modeling with the inclusion of π-stacking restraints at the overlapped drug region would be required in order to confirm these hypotheses.

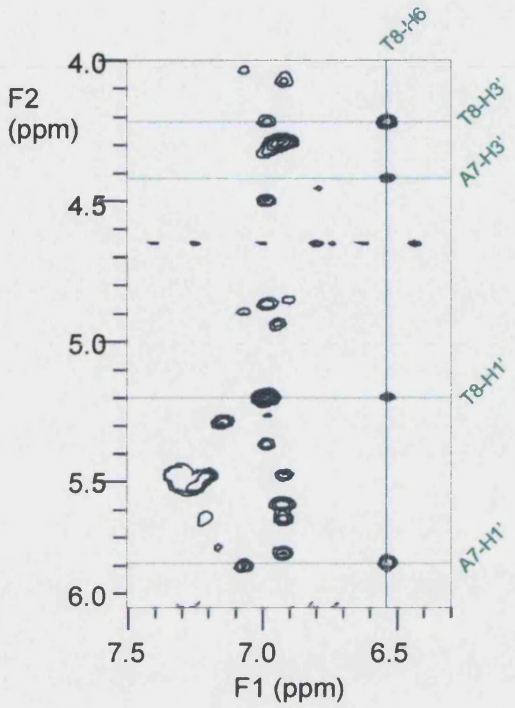


Figure 122: 600 MHz 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin adduct, 200 ms mixing time. Expansion showing T8-H6 Watson-Crick resonances, it can be observed that the T8-H6 – H1' cross-peaks are very weak relative to other H1' peaks

The 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct displays a perhaps surprising lack of changes in chemical shift values at the central region of the duplex in the area of drug overlap. These central protons remain at a chemical shift very close to that found in the duplex. The ability of the minor groove to open by up to 17 Å has been noted in the studies of lexitropsins.[171] These are non-covalent minor groove binders that form overlapped adducts, and are discussed in **Chapter 1**. Presumably on widening in this manner the steric crowding and subsequent perturbation of the DNA backbone in this area is reduced sufficiently that the chemical shifts of these protons remain relatively unchanged.

	H6/H8			H1'			CH_3/H5/H2			H2'			H2''			H3'			H4'			Imino
C1	7.40	7.40	0.00	5.47	5.99	0.52	5.67	5.67	0.00	1.65	2.19	0.54	2.16	2.45	0.29	4.51	4.51	0.00	3.52	3.80	0.28	*
G2	7.80	7.76	0.02	5.36	5.29	0.07	-	-	-	2.55	2.55	0.00	2.66	2.64	0.02	4.83	4.14	0.69	4.17	3.04	0.13	-
A3	8.10	8.11	0.01	6.13	6.13	0.00	7.40	7.40	0.00	2.54	2.57	0.03	2.78	2.78	0.00	4.89	4.89	0.00	4.00	3.90	0.10	-
T4	7.00	7.05	0.05	5.71	5.79	0.08	1.17	1.11	0.06	1.79	1.77	0.02	2.32	2.34	0.02	4.71**	4.03	-	4.00	3.49	0.51	12.80
T5	7.16	7.29	0.13	5.52	5.34	0.18	1.43	1.49	0.06	1.93	2.09	0.16	2.28	2.41	0.13	4.71**	4.21	-	3.96	3.31	0.65	9.76
A6	8.10	8.22	0.12	5.76	5.90	0.14	*	7.56	-	2.58	2.53	0.05	2.74	2.66	0.08	4.33	4.85	0.52	4.28	3.76	0.52	-
A7	7.97	7.81	0.16	5.96	5.78	0.18	7.65	7.82	0.17	2.36	2.09	0.27	2.71	2.41	0.30	4.82	4.40	0.42	4.00	3.00	1.00	-
T8	6.94	6.51	0.43	5.70	5.18	0.52	1.09	0.97	0.12	1.78	1.34	0.44	2.22	1.84	0.38	4.90	4.21	0.69	3.96	4.00	0.04	11.02
C9	7.28	7.96	0.68	5.53	5.38	0.15	5.43	5.18	0.25	1.83	1.49	0.34	2.18	1.89	0.29	4.71**	4.29	-	3.90	3.80	0.10	*
G10	7.75	7.56	0.19	5.94	5.73	0.21	-	-	-	2.18	2.07	0.11	2.44	2.31	0.13	4.51	4.34	0.17	3.90	3.51	0.39	-

Table 23: Chemical shifts (ppm) for the 5'*d*(CGATTAATCG)$_2$ duplex and the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct, DNA resonances. Red = Duplex, Black = Adduct, Blue = difference, greater than 0.25 ppm underlined. * = peak not found ** =peak under residual water peak at 4.71 ppm

CH_3	H1A	H1B	H3	H6	H8A	H8B	H8A2	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
2.70	3.51	3.51	8.14	8.28	5.50	5.50	4.20	7.92	6.96	8.10	8.14	7.28	7.80	7.46	7.53	8.10

Table 24: Chemical shifts (ppm) for the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct, drug peaks

5.6: Intra-molecular drug/DNA contacts – Watson-Crick

The assignment of the complete 2D NOESY and COSY spectra for the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct provided many drug-DNA inter-proton connectivities to confirm the location of the drug in the minor groove of the DNA duplex. A table showing all drug-DNA connectivities can be found in **Table 25,** while a diagram depicting these connectivities is shown in **Figure 124.**

NOE cross-peaks between the DNA duplex and the 'A' subunit of adozelesin confirm the association of the drug molecule with the base protons at the site of covalent modification. In particular an intense NOE between A3-H2 and Ado-H1a/b provides strong evidence for the covalent modification site at A3. (**Figure 123**)

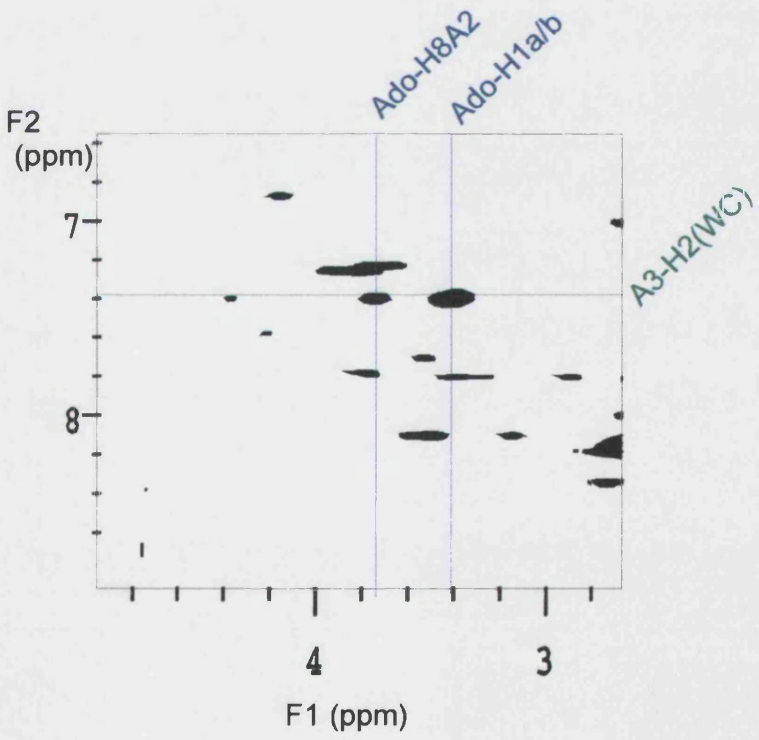


Figure 123: 600 MHz 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct 300 ms mixing time. Expansion showing NOE cross-peaks between A3-H2(WC) and the CPI head unit confirming covalent link site

5.6.1: Association of the Watson-Crick adduct with the modified strand of DNA.

NOE connectivities between the protons of the 'C' subunit (Ado-H4'2, Ado-H5'2, Ado-H6'2 and Ado-H7'2) and bases A6, A7 and T8 as shown in **Table 25** confirm a close association of the drug with its own modified strand as shown in **Figure 121 B (section 5.4.1).** If the 'C' subunits were overlapped in such a way as to afford a closer association with the opposite strand (as in **Figure 121 A**) then these protons would be too spatially distant to exhibit NOE connectivities. **Figure 125** shows an expansion of the 2D NOESY H6/H8-H1' region, with cross-peaks from aromatic drug protons to DNA H1' protons marked. A pictorial representation of this association can be viewed in the refined molecular model of the adduct in **section 5.7**.

5.6.2: Overlap of the 'C' subunits results in dipolar coupling of 'B' and C' subunit protons

NOE resonances between the Ado-H6'1 and Ado-H3'1 protons of the indole 'B' subunit and the aromatic protons of the benzofuran 'C' subunit such as Ado-H6'1 to Ado-H4'2, and Ado-H6'1 to Ado-H6'2 (**Figure 126**) confirm the overlap of the 'B' and 'C' subunits. These connectivities can only arise from the overlap of the 'B' and 'C' subunits in the respective molecules of adozelesin in the *bis*-adduct, as the protons are too spatially separated from each other to exhibit connectivities within the same drug molecule. These observations provide evidence that the subunits are substantially overlapped, as depicted in **Figure 108** in **section 5.1**. Once again, a pictorial representation can be viewed on observation of the refined molecular model in **section 5.7**.

	H1a/b	H8a/b	H8A2	CH_3	H6	H3	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
A3-H8				O											
A3-H1'				O											
A3-H4'				O											
A3-H2	S	M(O)					W(O)								
T4-H1'		O				VW	M/W								
T4-H3															
T5-H1'												M(O)			
T5-H3'													W		
T5-H4'								W							
A6-H1'			M								W(O)	O			
A6-H8											W				
A6-H4'															
A7-H2	M		M/W				O	W				O	O	M	
A7-H1'								W			M(O)	W(O)	M(O)	W	
A7-H3'								W							M(O)
T8-H1'										VW					
T8-H4'									W					W	

Table 25: NOE connectivities between the drug protons and the DNA duplex in the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct. S=Strong, M= Moderate, W= Weak, VW = Very weak, O = Overlaid

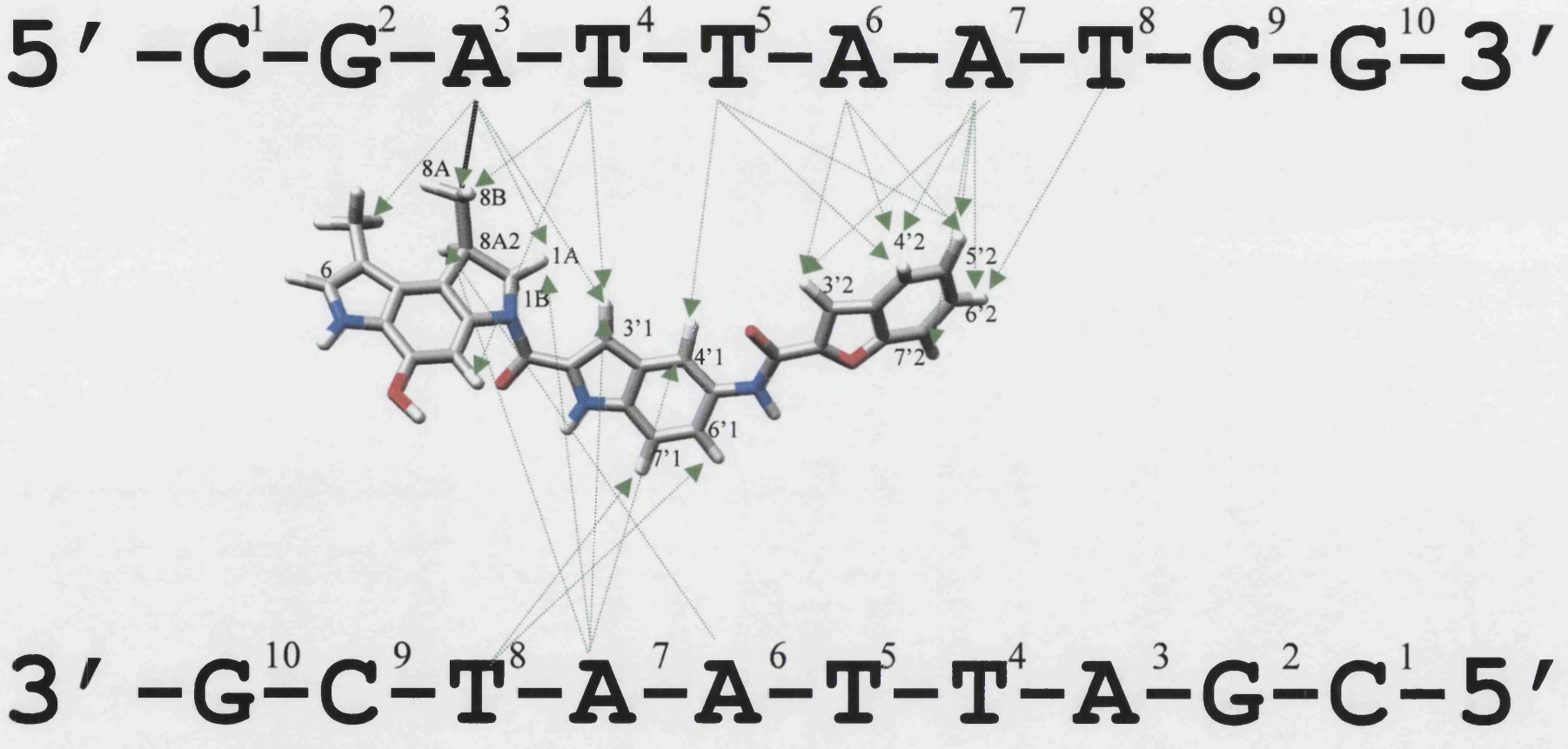


Figure 124: Representation of one end of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct, showing NOE connectivities between the drug and the DNA backbone

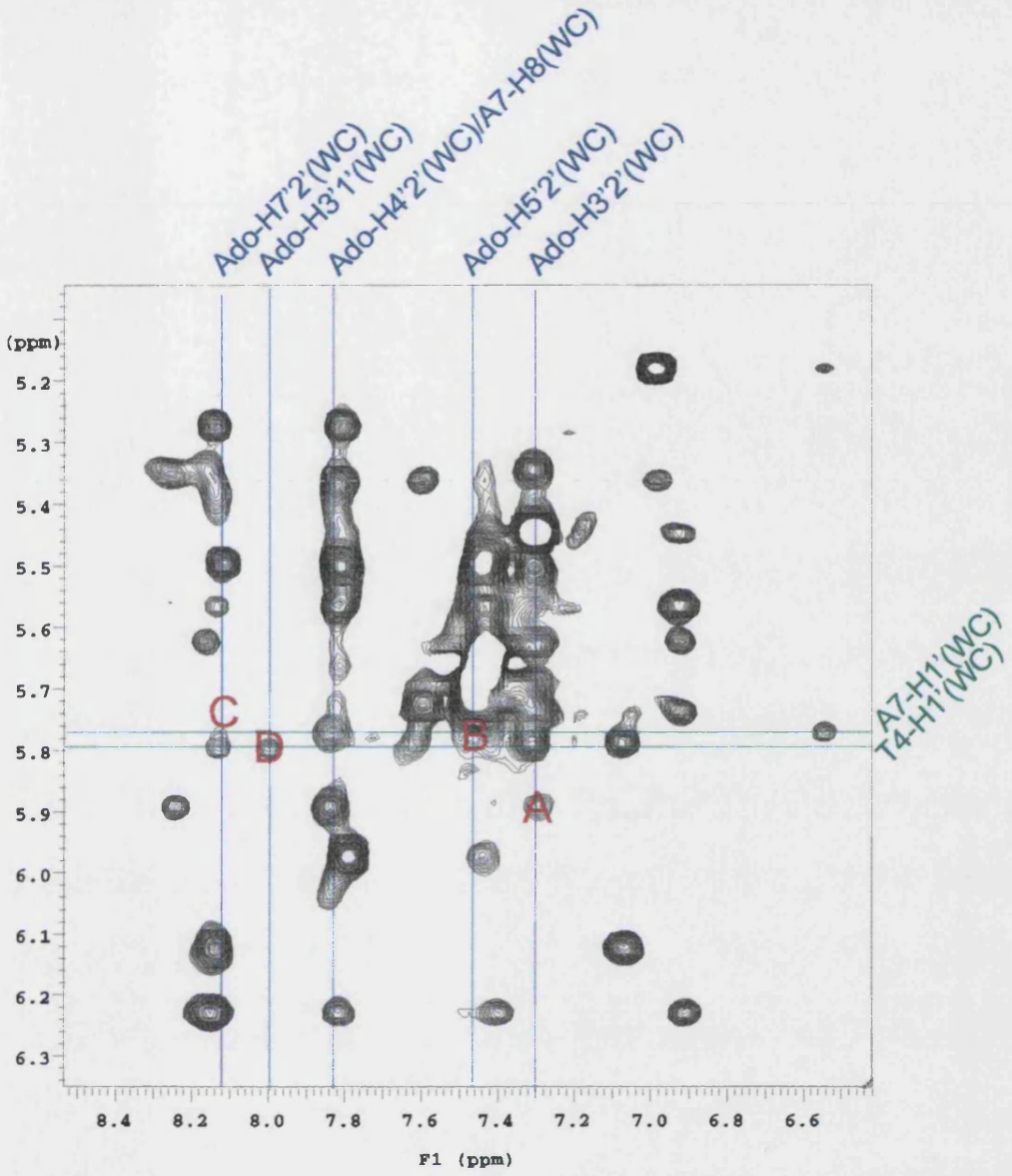


Figure 125: 600 MHz 2D NOESY spectrum of 5'*d*(CGATTAATCG)$_2$-adozelesin, mixing time 200ms. Expansion of the H6/H8-H1' region . Peaks A, B, C, and D are cross-peaks that have arisen as a result of the association of the drug 'C' subunit with its own covalently modified strand

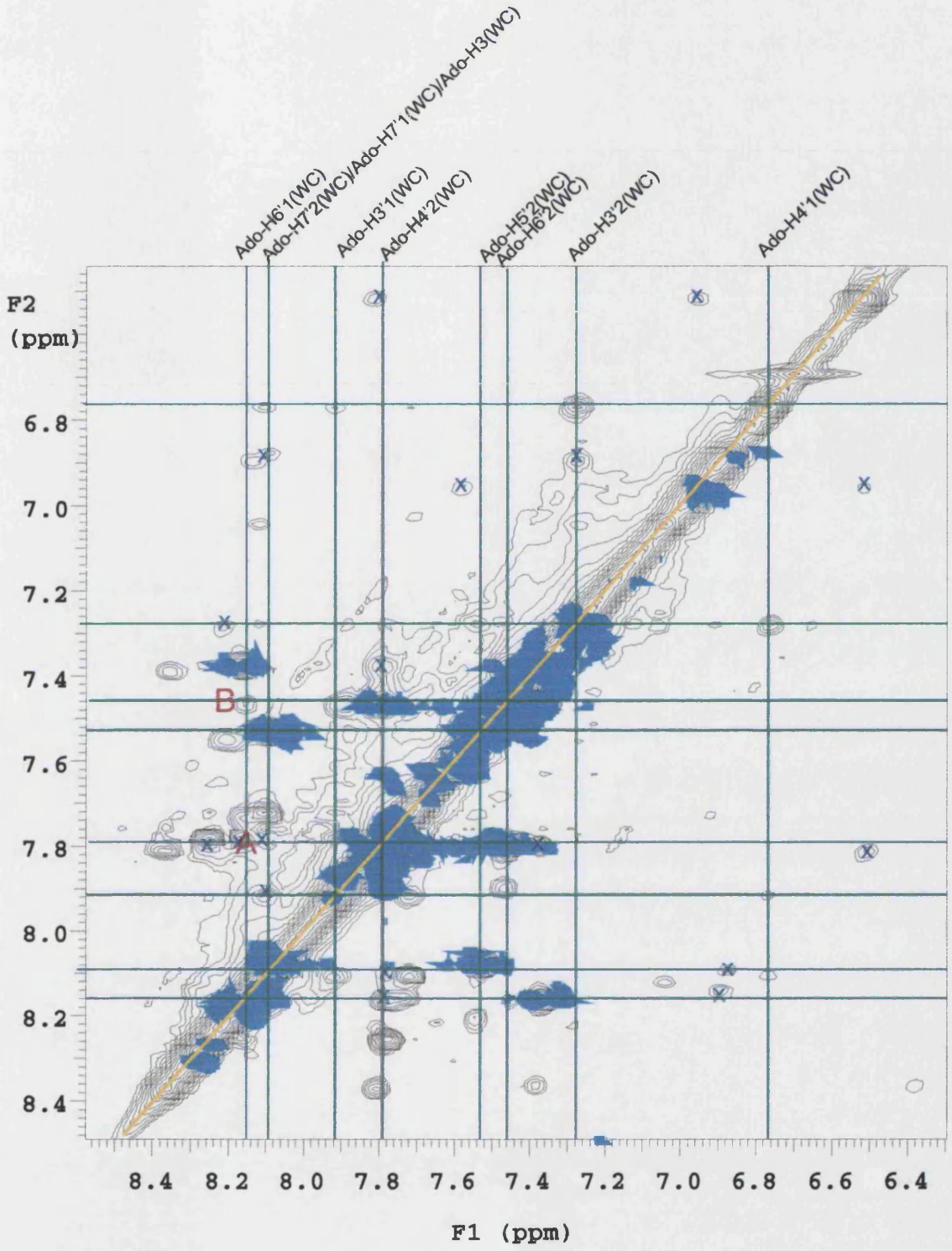


Figure 126: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. NOESY mixing time 200ms. Black=NOESY, Blue=COSY. Expansion of the H6/H8-H6/H8 region. Peaks A and B show connectivities arising from the overlap of the 'C' subunits. X = peaks due to sequential connectivities of the DNA backbone

5.7: The refined molecular model of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct

The 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct was modelled using the SYBYL software suite[197] as described in **Chapters 3** and **4**. The drug-DNA adduct was modelled initially within HYPERCHEM[198] and subjected to steepest descent energy minimization and molecular dynamics *in vacuo* at 300 K. On completion the overlapped adozelesin molecules were extracted as a pair and docked onto a Watson-Crick duplex produced within the SYBYL package. The model was then subjected to procedures identical to those discussed in **Chapter 3**. Complete restraints tables can be seen in **Appendix III**.

Figures 127 & 128 show the *in aquo* computer model of the adozelesin Watson-Crick adduct. Adozelesin is shown in white, with the two DNA strands in green and purple. As with previous adducts it is clear that Watson-Crick base pairing has been maintained, with retention of the β-helical structure that allows sequential assignment of the DNA duplex.

Figure 129 shows the central region of the duplex, with A7-H1', Ado-H3'2 and Ado-H7'2 highlighted. NOE connectivities between the adozelesin aromatic protons and A7-H1' are a clear indication of the drug association with its own covalently modified strand as discussed in **section 5.6.1** as any other configuration would render these protons too spatially separated to exhibit NOE connectivities in this manner.

Figure 130 again shows the central region of the duplex. Ado-H4'2, Ado-H6'2 Ado-H4'2 and Ado-H6'1 are highlighted in red. These protons were found to be dipolar coupled, exhibiting NOESY cross-peaks as discussed in **section 5.6.2**. This provides compelling evidence for the overlap of the 'C' subunits.

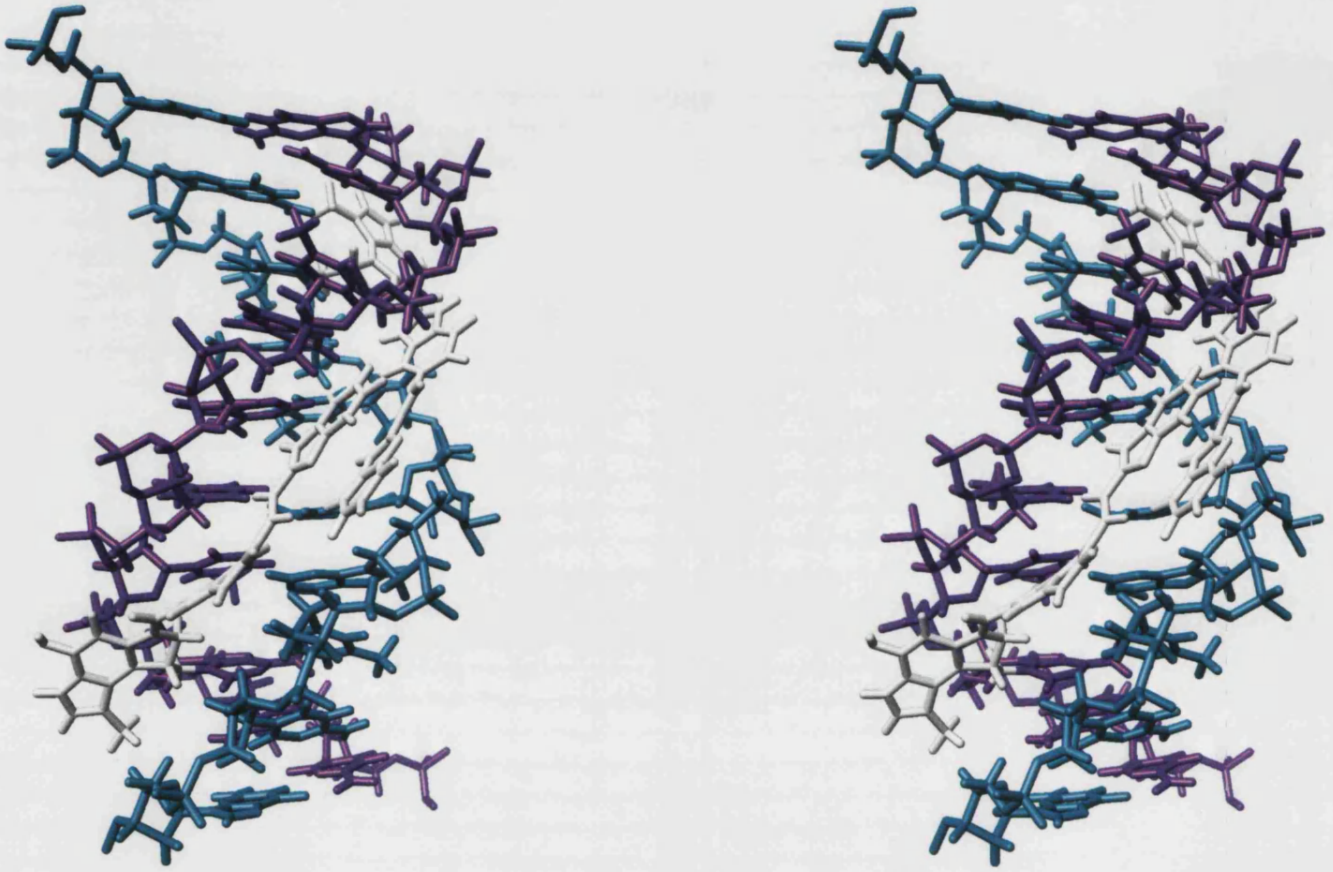


Figure 127: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct. DNA strands are shown in purple and green while the adozelesin molecules are shown in white. The duplex has maintained β-helical structure and the adozelesin molecules are accommodated snugly following the contour of the minor groove. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

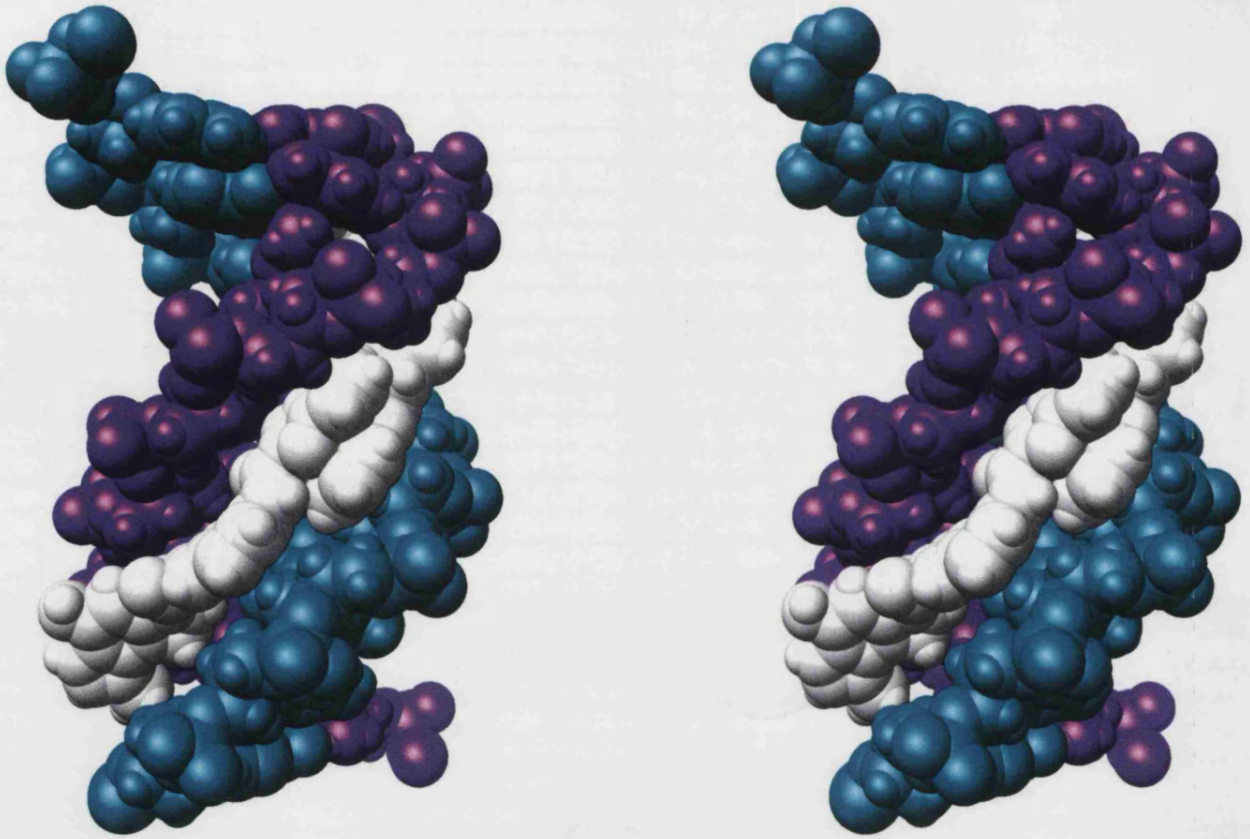


Figure 128:_Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct. DNA strands are shown in purple and green while the adozelesin molecules are shown in white. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

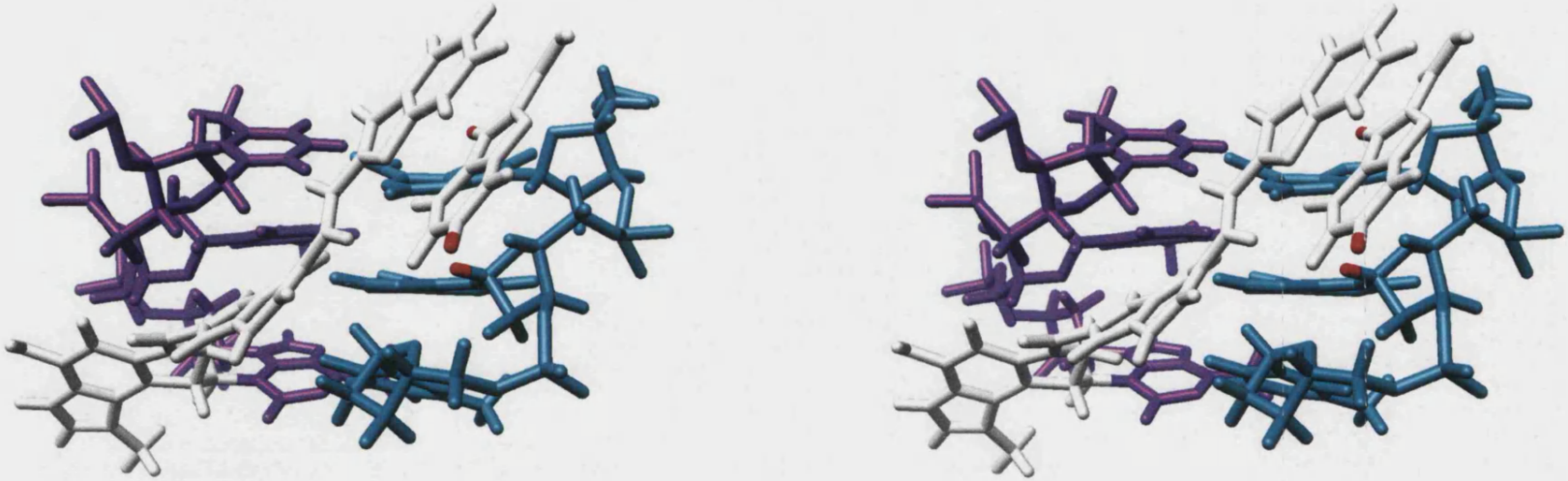


Figure 129: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct, central region. A7-H1', Ado-H3'2 and Ado-H7'2 are shown in red. NOE connectivities between these protons confirm association of the adozelesin molecule with its own covalently modified DNA strand as pictured. Models produced in SYBYL software suite[197] and pictured using UCSF Chimera[193]

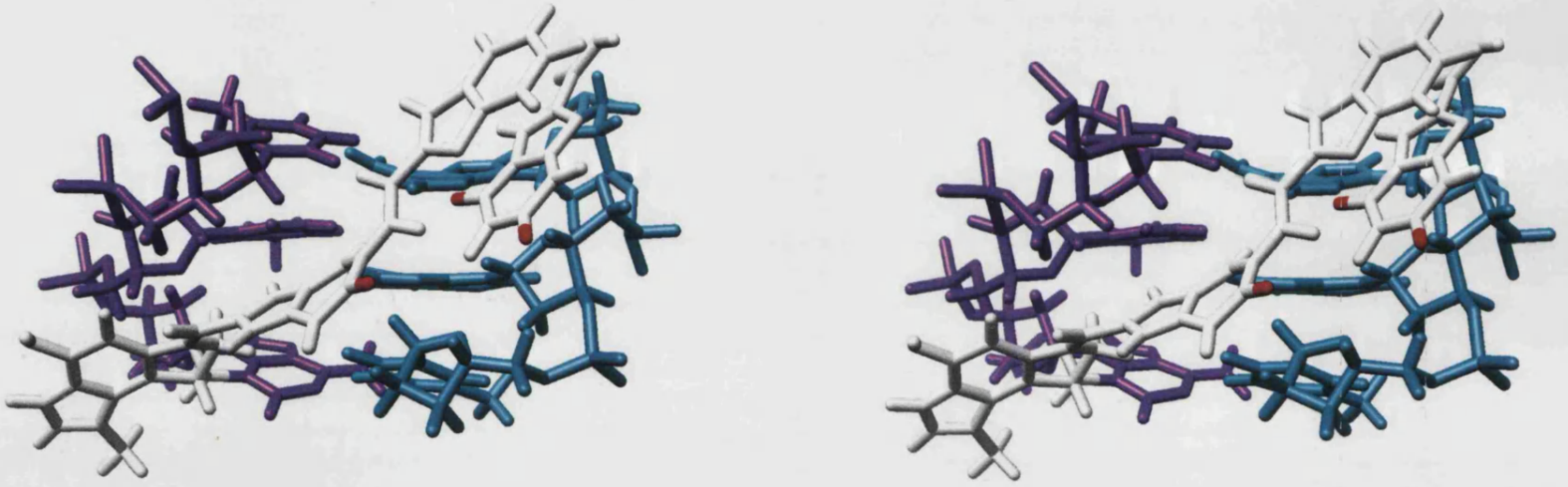


Figure 130: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct, central region. Ado-H6'1, Ado-H4'2 andAdo-H6'2 are shown in red. NOE connectivities between these protons confirm overlap of the 'C' subunits as pictured.Models produced in SYBYL software suite[197] and pictured using UCSF Chimera[193]

5.8: Nucleic acid proton chemical shifts of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct

Proton chemical shifts in the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct relative to duplex DNA can be found in **Table 26**. As found in the Watson-Crick adduct, down-field shifts of the adozelesin cyclopropyl ring protons Ado-H8A and Ado-H8B from the expected resonance of 1.47 ppm and 2.02 ppm respectively[200] to 5.35 ppm confirms the site of covalent attachment as Ado-C8 as discussed in **section 5.3**. As expected on accommodation of the adozelesin molecule in the minor groove some large shift changes can be observed.

A large up-field shift of 0.50 ppm can be observed in the H1' proton of the A3 residue. As discussed in **section 5.3** this is the site of covalent modification and as such large shifts here are not unexpected. A significant up-field shift (0.45 ppm) was also seen in the 5'*d*(CGTAAGCGCTTACG)$_2$-adozelesin adduct studied previously at the equivalent A12 site.[104]

Large changes in chemical shift of the T5-H8 proton as well as the A6-H1' and A7-H1' protons were also observed in the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct. The T5 and A6 nucleotides are the sites of the Hoogsteen base pair, and the large chemical shift changes lend support to a change of environment at this point as a result. In addition these central protons are all situated in the vicinity of the overlapping 'C' subunits of the adozelesin molecules so large changes in chemical shift could be indicative of some steric crowding at this point.

5.9: Intra-molecular drug/DNA contacts – Hoogsteen

Once again, careful assignment of the 2D NOESY and COSY spectra for the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct provided many inter-proton connectivities from drug to DNA, confirming the location of the drug in the minor groove. A table showing all drug-DNA connectivities can be found in **Table 28** whilst a diagram depicting these connectivities is seen in **Figure 131.**

	H6/H8			H1'			CH_3/H5/H2			H2'			H2''			H3'			H4'			Imino
C 1	7.40	7.41	*0.01*	5.47	5.51	*0.04*	5.67	5.68	*0.01*	1.65	1.63	*0.02*	2.16	2.22	*0.06*	4.51	4.69	*0.20*	3.52	3.89	*0.37*	*
G 2	7.80	7.79	*0.01*	5.36	5.34	*0.02*	-	-	-	2.55	2.60	*0.05*	2.66	2.70	*0.04*	4.83	4.23	*0.60*	4.17	*	-	-
A 3	8.10	8.15	*0.05*	6.13	5.63	*0.50*	7.40	7.28	*0.12*	2.54	2.24	*0.30*	2.78	2.62	*0.16*	4.89	4.97	*0.08*	4.00	3.91	*0.09*	-
T 4	7.00	6.90	*0.10*	5.71	5.58	*0.13*	1.17	1.16	*0.01*	1.79	1.25	*0.49*	2.32	1.40	*0.08*	4.71**	4.91	*0.20*	4.00	3.50	*0.50*	10.13/ 11.07
T 5	7.16	6.92	*0.24*	5.52	5.58	*0.06*	1.43	1.32	*0.11*	1.93	1.62	*0.31*	2.28	1.83	*0.45*	4.71**	4.28	*0.43*	3.96	3.32	*0.64*	12.90
A 6	8.10	7.80	*0.30*	5.76	5.39	*0.37*	*	7.43	-	2.58	2.57	*0.01*	2.74	2.70	*0.04*	4.33	4.88	*0.55*	4.28	3.80	*0.52*	-
A 7	7.97	8.14	*0.17*	5.96	6.23	*0.27*	7.65	7.82	*0.17*	2.36	2.51	*0.15*	2.71	2.59	*0.12*	4.82	4.40	*0.42*	4.00	3.32	*0.68*	-
T 8	6.94	6.90	*0.04*	5.70	5.74	*0.04*	1.09	1.13	*0.04*	1.78	2.13	*0.35*	2.22	2.20	*0.02*	4.90	4.09	*0.81*	3.96	4.01	*0.05*	10.13/ 11.17
C 9	7.28	7.29	*0.01*	5.53	5.51	*0.02*	5.43	5.45	*0.02*	1.83	1.84	*0.01*	2.18	2.22	*0.04*	4.71**	3.96	*0.75*	3.90	3.89	*0.01*	*
G 10	7.75	7.41	*0.34*	5.94	5.68	*0.26*	-	-	-	2.18	1.68	*0.50*	2.44	2.09	*0.35*	4.51	4.40	*0.11*	3.90	3.55	*0.35*	-

Table 26: Chemical shifts (ppm) for the 5'*d*(CGATTAATCG)$_2$ duplex and the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct. Red = duplex, Black = adduct, Blue = difference, greater than ±0.25 ppm underlined. * = not found ** = situated under residual water peak at 4.71 ppm

CH_3	H1A	H1B	H3	H6	H8A	H8B	H8A2	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
2.83	3.89	3.89	8.10	8.28	5.35	5.35	3.90	7.28	7.80	8.37	8.34	7.72	8.11	8.16	7.38	7.45

Table 27: Chemical shifts (ppm) for the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct, drug peaks

Numerous connectivities between the drug and the 'A' subunit of adozelesin once again confirm the association of the drug molecule with the base protons surrounding the A3 nucleotide. As expected, connectivities from the 'A' subunit are predominantly to bases A3 and T4 of the DNA duplex, owing to the close association of the head unit with the modified strand on formation of a covalent link.

5.9.1: NOE connectivities confirm an overlap of the 'C' subunits

Evidence for the overlap of the 'B' and 'C' subunits exists in the form of connectivities between the subunits. NOE connectivities can be seen between the Ado-H4'1 proton of the indole ('B') subunit and the aromatic Ado-H4'2, Ado-H5'2 and Ado-H6'2 protons of the benzofuran ('C') subunit. In addition a cross-peak can be seen as a result of the spatial proximity of Ado-H6'2 to Ado-H6'1. **(See Figure 132).** As discussed in **section 5.6.2,** these connectivities can only arise from the overlap of these subunits in the respective molecules of adozelesin. Such cross-peaks are a firm indication of a substantial overlap of the 'C' subunits. A pictorial representation of the spatial separation of these protons can be found in the refined molecular model of the adozelesin-Hoogsteen adduct in **section 5.10.**

5.9.2: NOE connectivities confirm the association of the adozelesin molecule with the covalently modified DNA strand.

When the NOE connectivities between the drug and the DNA backbone are examined it is evident that there are cross-peaks between the A7-H1' base proton of the DNA backbone and the aromatic Ado-H5'2 and Ado-H6'2 protons of the adozelesin benzofuran subunit (**see Figure 133**). As discussed in **section 5.6.1** such dipolar coupling of adozelesin aromatic protons with the base protons of the covalently modified strand provides unequivocal evidence for the close association of the drug molecule with the modified strand as depicted in **Figure 121 B.**

	H1a/b	H8a/b	H8A2	CH_3	H6	H3	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
A3-H8				VW											
A3-H1'		O	VW												
A3-H4'		M/W		W	W										
A3-H2	M	W(O)	M/W												
T4-H1'	O					W	M(O)								
T4-H3								W(O)							
T4-H4'			O												
T5-H1'												W(O)	W(O)		
T5-H3'								W					W		
T5-H4'								W					M		
T5-H6							M(O)					O			
A6-H1'								O			W	O			
A6-H8								O			W(O)				
A6-H4'								W(O)							
A7-H2	O						(O)	O				O	M(O)		
A7-H1'								M			M(O)	W(O)	M	M	VW
A7-H3'								W(O)							M(O)
A7-H4'												M/W		W(O)	W
T8-H1'	VW(O)						M(O)			VW					
T8-H6													O		

Table 28: NOE connectivities between the drug protons and the DNA duplex in the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct. S=Strong, M= Moderate, W= Weak, VW = Very weak, O = Overlaid

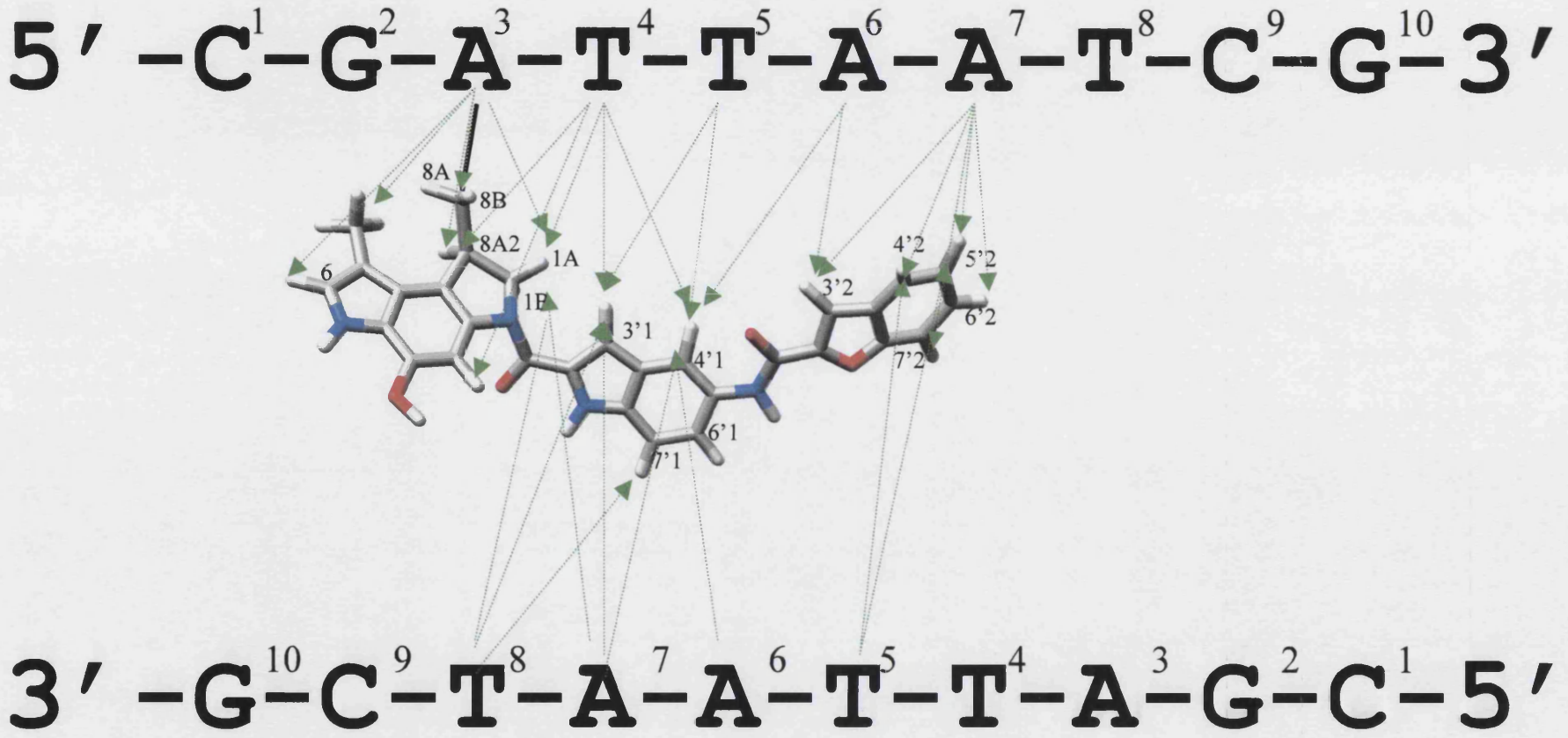


Figure 131:Representation of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct, showing NOE connectivities between the drug and the DNA duplex. One adozelesin molecule is shown only to aid clarity

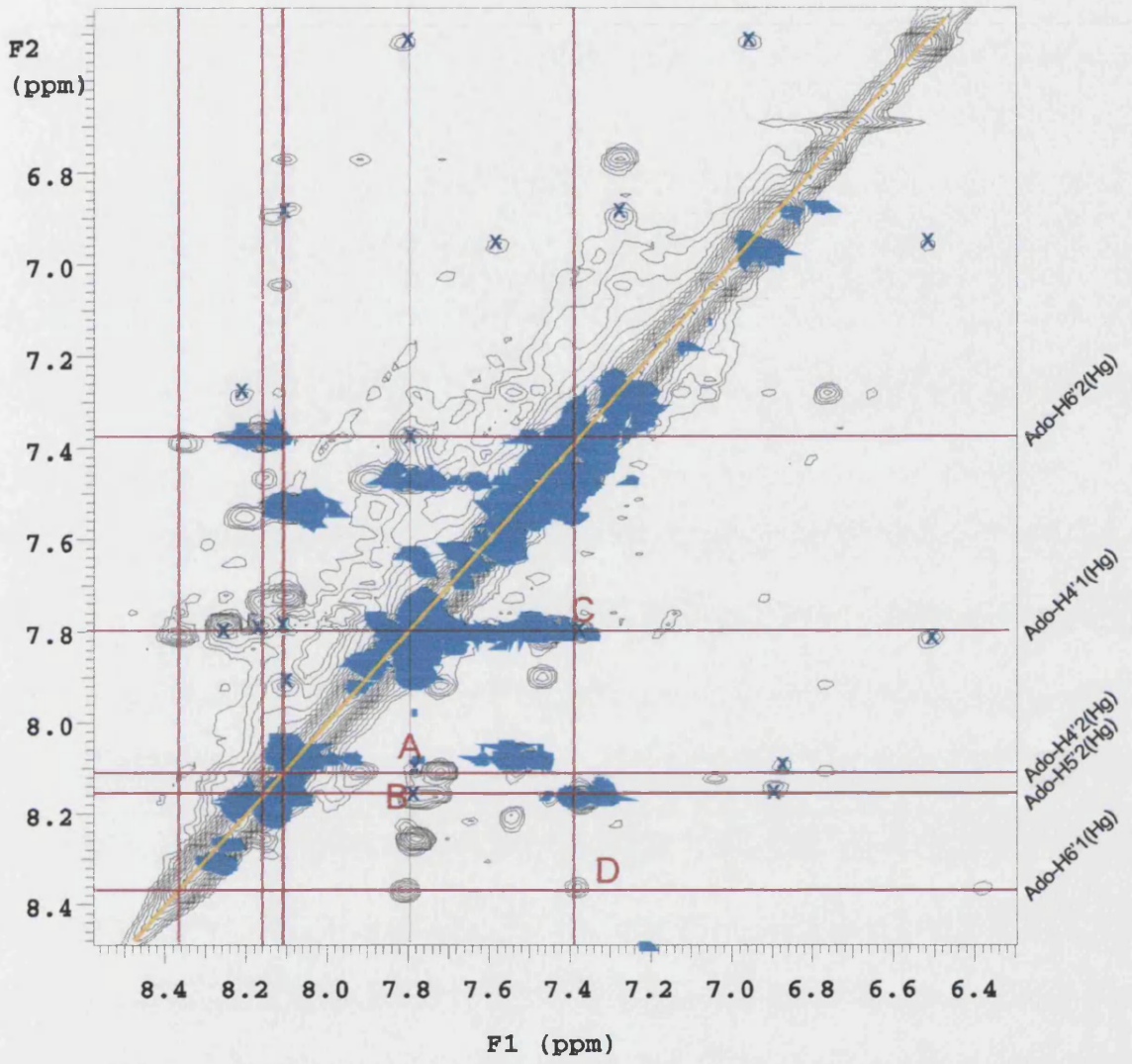


Figure 132: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct, NOESY 200 ms mixing time. Black=NOESY, blue=COSY. Expansion of the H6/H8-H6/H8 region. Peaks A, B, C and D show NOEs arising as a result of the 'C' subunit overlap, other assignments are omitted to aid clarity. X = peaks due to sequential connectivities of the DNA backbone

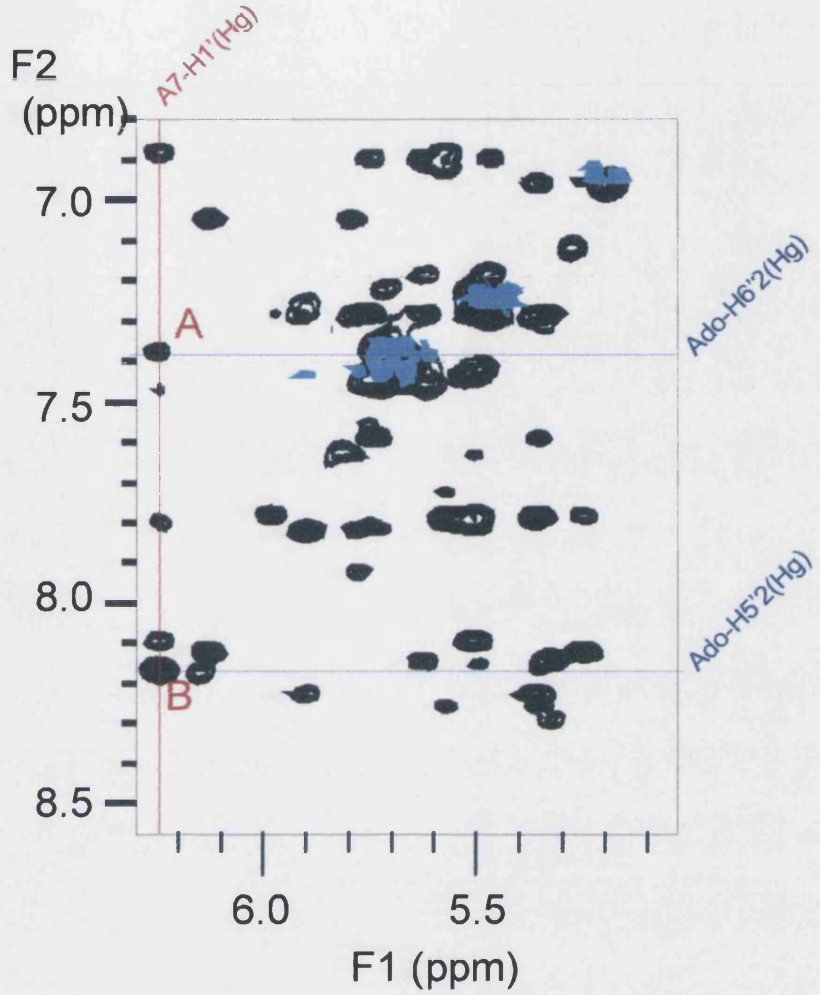
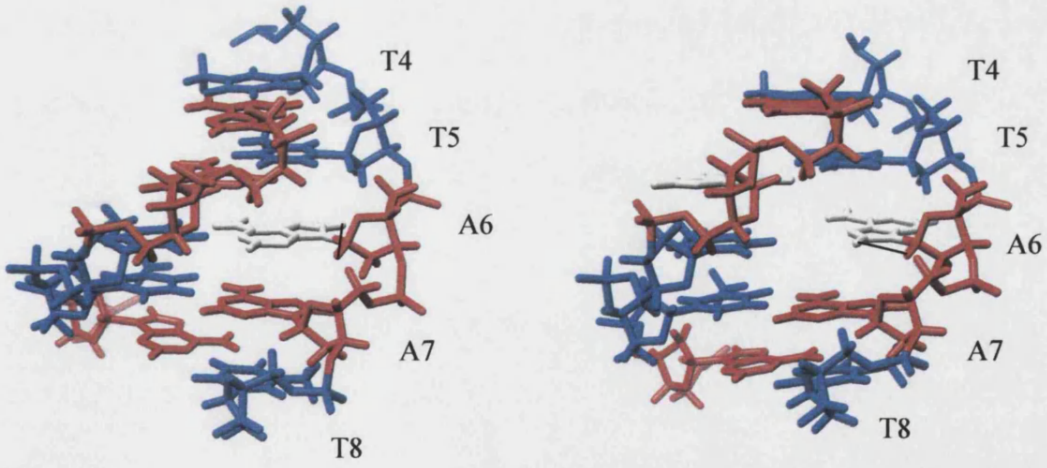


Figure 133: 600 MHz 2D NOESY and COSY spectra of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. NOESY mixing time 200 ms, expansion of the H6/H8-H1' region. Black NOESY, Blue=COSY. Peaks A and B are NOE connectivities arising from the association of the 'C' subunit with the covalently modified strand. Other assignments are omitted to aid clarity.

5.9.3: Formation of Hoogsteen base pairs in the central region of the 5'd(CGATTAATCG)$_2$-adozelesin adduct

The most diagnostic cross-peaks for a Hoogsteen base pair come from A6-H2 and A8-H8 protons, as seen in previous studies involving assignment of Hoogsteen duplexes.[115] In a Watson-Crick duplex the A-H2 is found in the minor groove while the major groove houses the A-H8 proton. When a Hoogsteen base pair is formed *anti* to *syn* rotation around the glycosidic bond (C1'-N9) moves the A-H2 to a major groove position similar to that previously occupied by A-H8 and relocates A-H8 in the minor groove. (**See Figure 119, section 5.4.1).** The new minor groove position of the *syn* A-H8 results in the classical connectivities to major groove protons such as DNA H3', H4', H2' and H2'' becoming greatly weakened, while the NOE cross-peak A6-H8 – A6-H1' is intensified due to the relatively shorter distance from the A6-H1' proton to the *syn* A6-H8. The resonances resulting from the interaction of A6-H8 with the protons discussed are shown in **Figure 135**, while a pictorial representation showing relative distances DNA-H1' to DNA *syn* A-H8 and DNA *anti* A-H8 can be seen in **Figure 134**.



Anti (Watson-Crick) A6-H8 *Syn* (Hoogsteen) A6-H8

Figure 134: Representation of the central region of the 5'*d*(CGATTAATCG)$_2$ duplex showing *syn* and *anti* A6-H8 (depicted in white). Lines show the A6-H8 – A6-H1' connectivity in each case

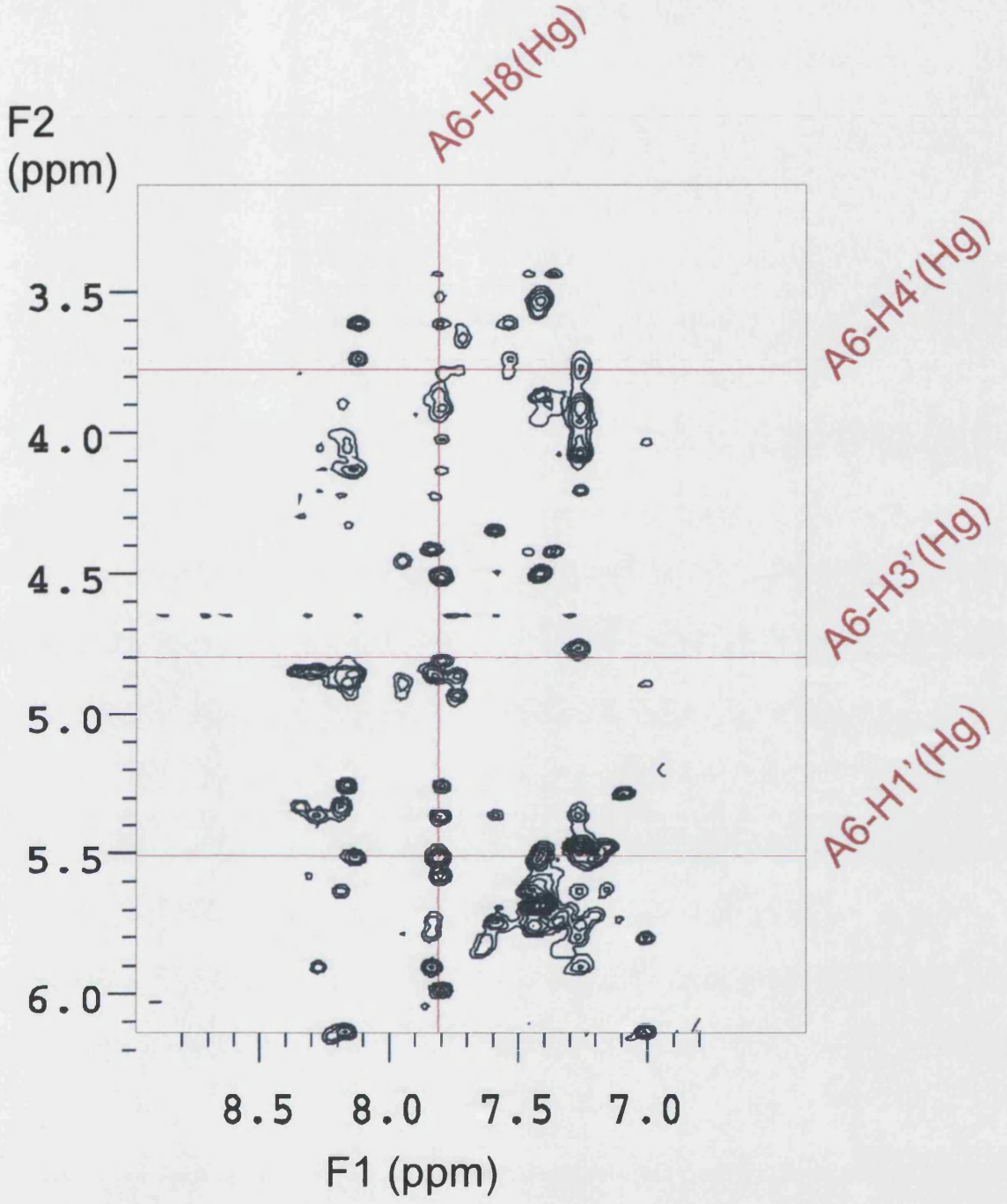


Figure 135: 600 MHz 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. 200 ms mixing time. Expansion showing NOE connectivities arising from the A6-H8 (Hg) resonance. Cross-peaks to protons in the minor groove are weakened while the cross-peak from A6-H8(Hg) to A6-H1'(Hg) is intensified

In addition to weakened peaks associated with the A6-H8 proton NOE resonances can be observed between the A6-H2 proton and the H2' and H2'' protons of the 3' neighbouring base (T5). As shown in **Figure 136** in its new location in the major groove the *syn* A6-H2 proton is placed in the vicinity of these protons and so exhibits NOE connectivities with them (**See Figure 137**). This is in analogy to previous studies on a Hoogsteen adduct performed by Seaman and Hurley.[115]

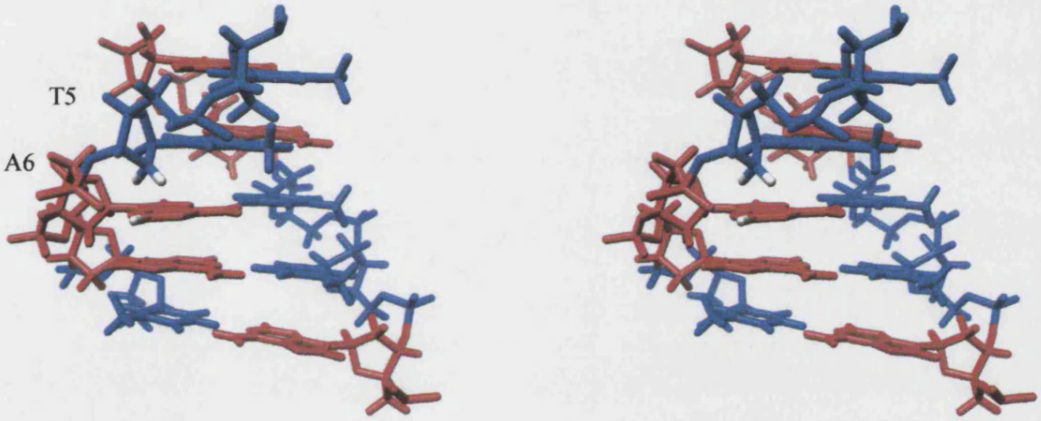


Figure 136: Stereoview of a DNA duplex showing a hoogsteen base at A6. The A6-H2 and T5-H2'/H2'' protons are shown in white

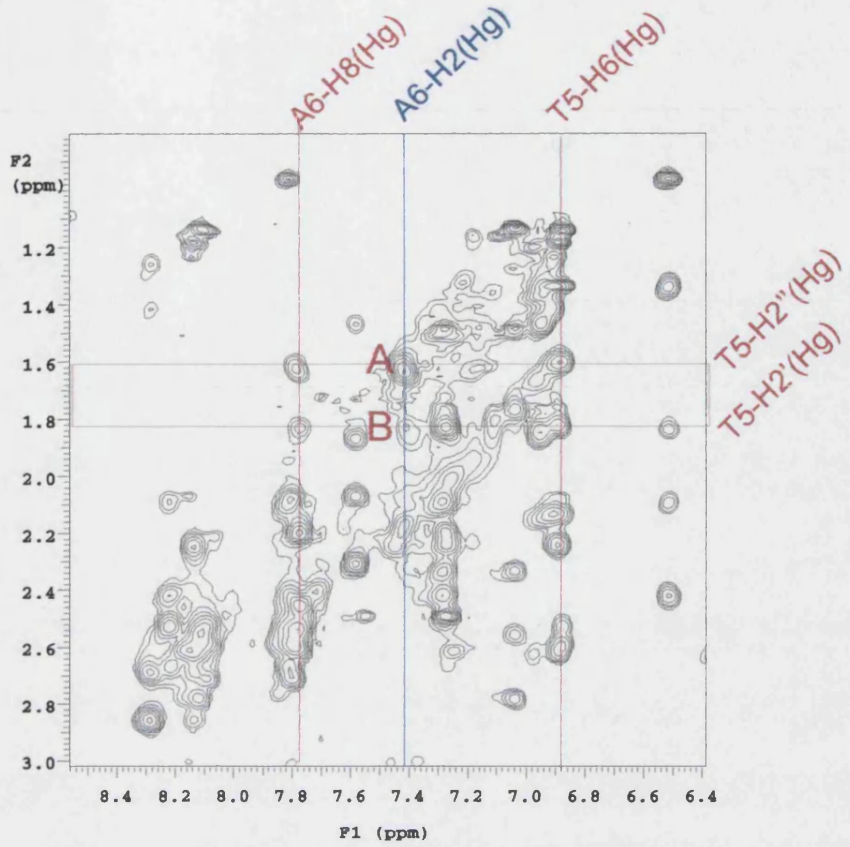


Figure 137: 600 MHz 2D NOESY spectrum of the 5'*d*(CGATTAATCG)$_2$-adozelesin mixed adduct. 200 ms mixing time. Expansion of the H6/H8-H2'/H2'' region showing NOE connectivity of A6-H2(Hg) with T5-H2'(Hg) and T5-H2''(Hg) (peaks A and B). Other assignments have been omitted to aid clarity.

5.10: The refined molecular model of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct.

The 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct was modelled using the SYBYL software suite[197] as described in **Chapters 3** and **4**. The Hoogsteen DNA duplex was produced within the SYBYL package and subjected to steepest descent energy minimisation prior to the docking of overlapped adozelesin molecules produced and energy-minimised using steepest descent and molecular dynamics at 300 K within HYPERCHEM.[193] The model was then subjected to procedures identical to those discussed in **Chapter 3**. Complete restraints tables can be seen in **Appendix IV.**

Figures 138 & 139 show the *in aquo* computer model of the adozelesin Hoogsteen adduct. Adozelesin is shown in white, with the two DNA strands in green and purple. With the exception of the central A6 Hoogsteen base pairs the duplex has retained β-helical structure. The central Hoogsteen pairs are slightly distorted but show minimal disruption to surrounding base pairs and accommodate the drug overlap by providing a widening of the minor groove.

Figure 140 shows the central region of the duplex, with A7-H1', Ado-H5'2 and Ado-H6'2 highlighted. NOE connectivities between the adozelesin aromatic protons and A7-H1' are a clear indication of the drug association with its own covalently modified strand as discussed in **section 5.6.1** as any other configuration would render these protons too spatially separated to exhibit NOE connectivities in this manner.

Figure 141 again shows the central region of the duplex. Ado-H4'2, Ado-H5'2, Ado-H6'2 and Ado-H6'1 are highlighted in red. These protons were found to be dipolar coupled, exhibiting NOESY cross-peaks as discussed in **section 5.6.2**. This provides compelling evidence for the overlap of the 'C' subunits as shown in **Figure 121 B.**

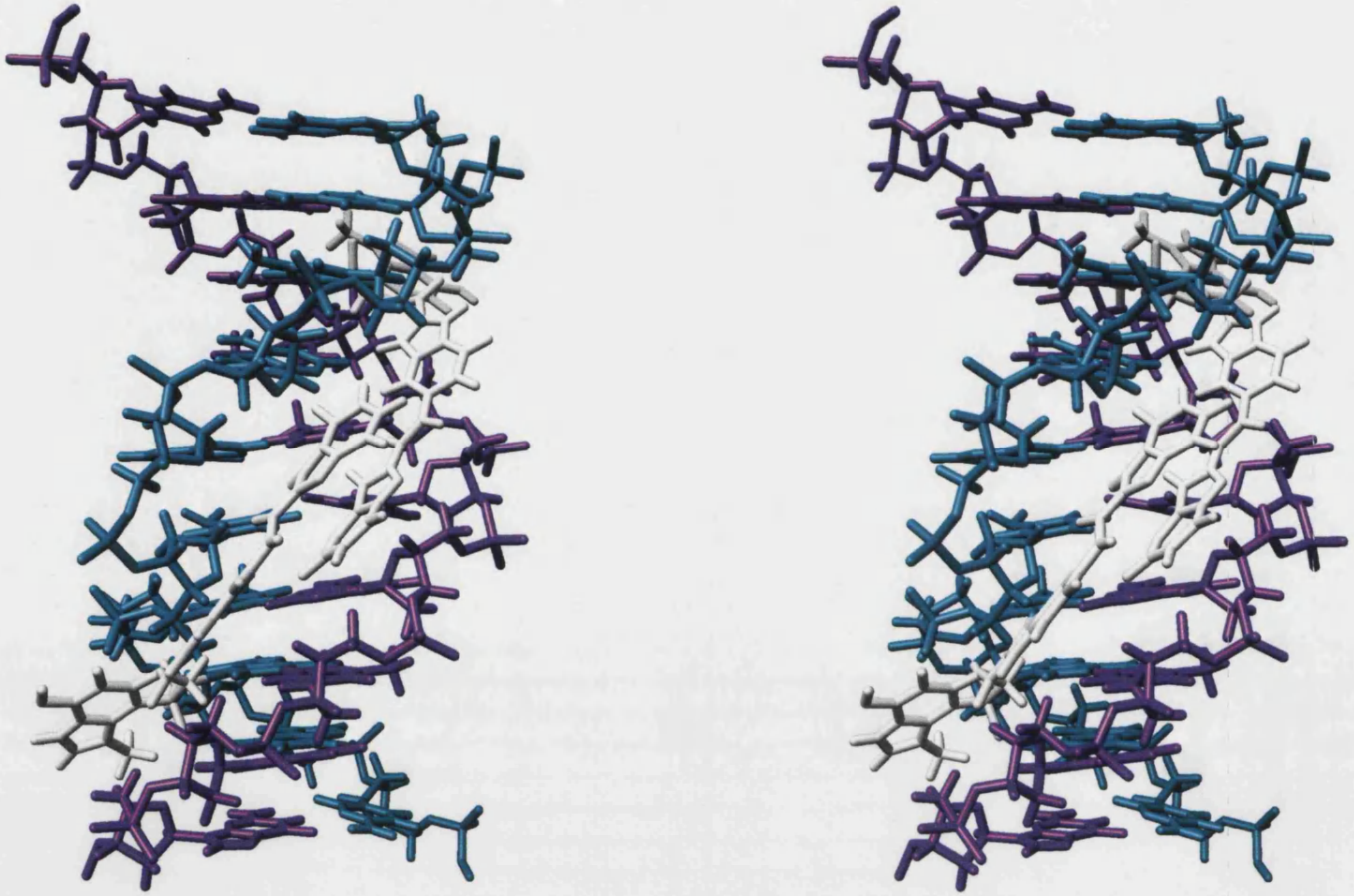


Figure 138: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct. Slight distortion can be seen at the central Hoogsteen base pairs, but the disruption to the surrounding bases is minimal. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

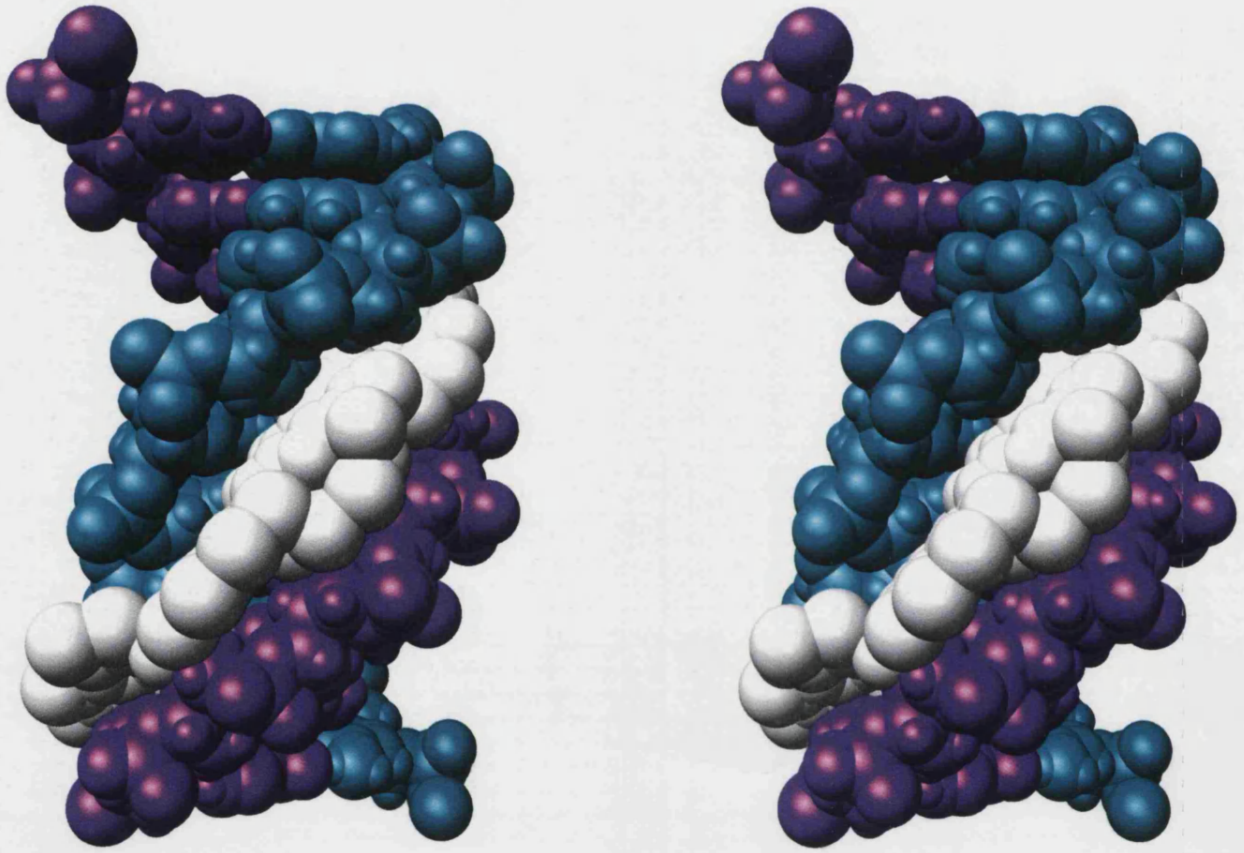


Figure 139: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct. The adozelesin molecule follows the contours of the minor groove with minimal disruption to the DNA helix. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

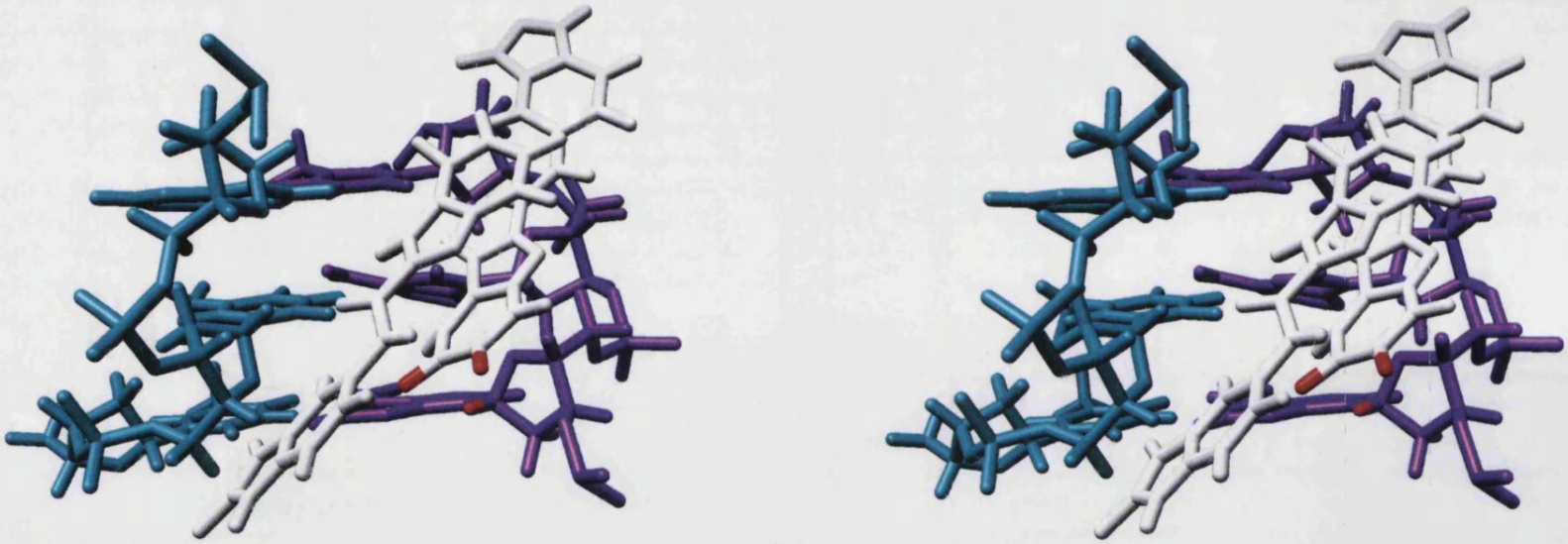


Figure 140: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct central region. A7-H1' and Ado-H5'2 and H6'2 are shown in red. A close association is observed resulting in NOE connectivities. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[195]

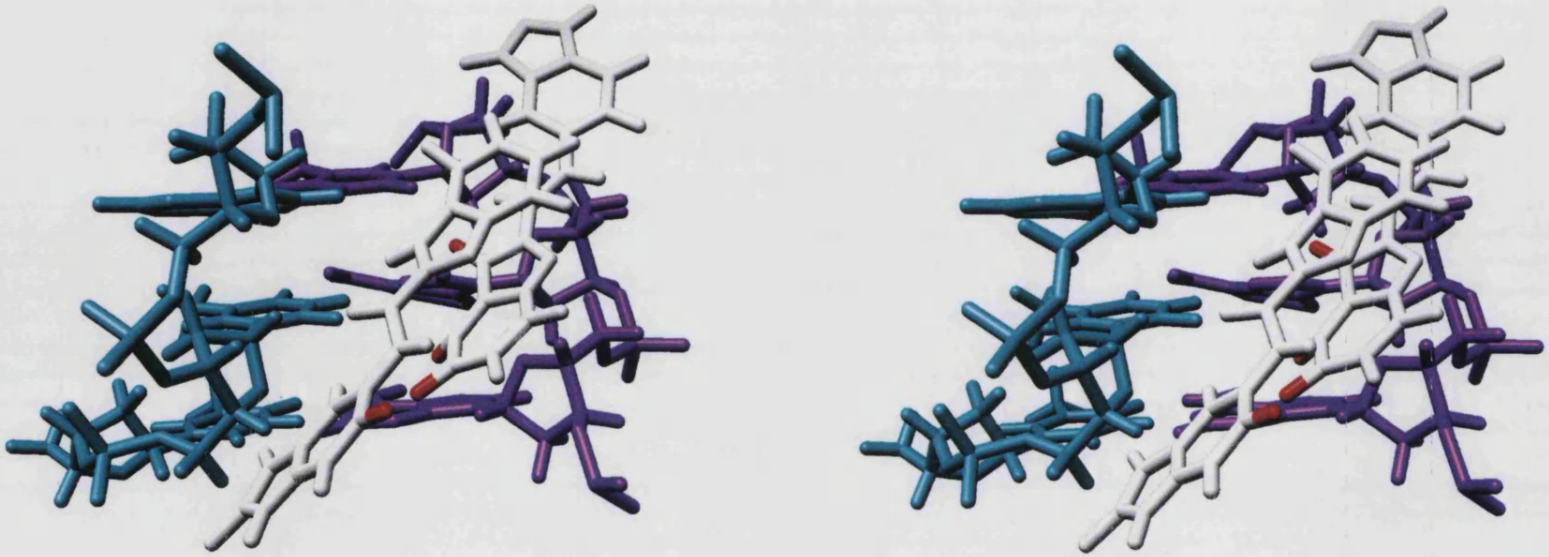


Figure 141: Stereoview of the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct central region. Ado-H4'2, Ado-H5'2, Ado-H6'2 and Ado-H6'1 are shown in red. Overlap of the 'C' subunit results in NOE connectivities. Models generated in the SYBYL software suite[197] and pictured using UCSF Chimera[193]

CHAPTER 6 – CONCLUSIONS

6.1: PBD Dimers

The covalent modification of two DNA sequences, 5'*d*(CICGATCICG)$_2$ and 5'*d*(CTCATCAC).(GTGATGAG), by SJG-136 has been investigated by high field NMR spectroscopy and a refined molecular model of each has been generated using distance-range restraints determined from the assignment of the 2D NOESY NMR spectra. The first adduct confirms the findings of previous studies of DSB-120 linkage with the same DNA sequence, while the second is the first example of a PBD-DNA intra-strand cross-linked adduct.

*6.1.1: The 5'*d*(CICGATCICG)$_2$-SJG-136 adduct*

As expected, an inter-strand cross-linked adduct was obtained from the reaction of SJG-136 within the minor groove of 5'*d*(CICGATCICG)$_2$. This result compliments very well studies by Mountzouris *et al* on the 5'*d*(CICGATCICG)$_2$-DSB-120 adduct.[181] The covalent reaction between the C11 carbon of SJG-136 and the exocyclic NH_2 group of the G4 nucleotide was confirmed. 'S' stereochemistry was assigned to C11, again in direct agreement with studies on the DSB-120 adduct, and providing support for previous molecular mechanics calculations, which predicted that the 'S' configuration was the energetically preferred stereochemistry in PBD-DNA adducts.[42,43,44]

Other similarities with the DSB-120 adduct include the reversal of the established chemical shift pattern of the DNA-H2' and H2'' protons. Typically, the H2' resonance is found up-field of the H2'' proton, located by the NOE cross-peak between DNA-H1' and DNA-H2'/H2''. In the case of both the DSB-120 adduct and the SJG-136 adduct this pattern is reversed and the more intense DNA-H1' – DNA-H2' coupling is found down-field of the DNA-H1' – DNA-H2''. This feature is indicative of conformational change within the internal nucleotide and confirms that, like DSB-120, SJG-136 induces additional perturbation of the DNA-C7 nucleotide, which faces the covalently modified G4 nucleotide in the minor groove. A further similarity between the DSB-120 and SJG-

136 adducts is the probable inversion of the aromatic rings of the drug, resulting in an attenuated up-field shift of the T6-H4' proton in each case, when compared with a DNA-tomamycin adduct.

As expected, differences between the SJG-136 and DSB-120 adducts can be mainly attributed to the presence of the ethylidene functionality in SJG-136. The shielding effect of this functional group causes up-field shifts of protons within the DNA-I8, DNA-C9 and DNA-C1 nucleotides. Interestingly, in the case of the SJG-136 adduct, evidence is present to support the location of SJG-136 as deeper within the minor groove than DSB-120, which was reported as shallowly immersed. This feature can also be attributed to the ethylidene group, with unsaturation at C2 leading to a better isohelical fit in the minor groove as discussed in **Chapter 1**.[73] In addition, there is no evidence for de-stacking of the I2-C3 step as observed with the DSB-120 adduct. It is possible that the superior fit of the SJG-136 molecule in the minor groove renders such perturbation of the DNA structure unnecessary in order to accommodate the drug.

6.1.2: The 5'd(CTCATCAC).(GTGATGAG)-SJG-136 adduct

The 5'*d*(CTCATCAC).(GTGATGAG)-SJG-136 adduct has been successfully made and represents a novel intra-strand cross-linking duplex adduct. The reaction sites are confirmed to be between the SJG-C11 centre and the *exo*cyclic NH_2 group of the G11 and G14 nucleotides. SJG-C11 stereochemistry is assigned as 'S' at both C11 reaction sites, in accordance with previous studies of PBD-DNA adducts both within this thesis (**Chapter 3**) and elsewhere,[45,181] and as expected the drug displays a greater association with the modified DNA strand.

As observed in both the DSB-120[181] and SJG-136 inter-strand adducts (**Chapter 3**), a reversal in the typical relative pattern of DNA-H2' and H2'' protons is noted, with the DNA-H2' proton resonating down-field of the DNA-H2'' proton. However, whilst in the inter-strand adducts this reversal was limited to nucleotides in the vicinity of the covalent modification site, the intra-strand adduct displays a pattern reversal throughout numerous nucleotides across the centre of the duplex, at nucleotides in the vicinity of the drug. This is attributed to some perturbation of the DNA structure on

accommodation of an intra-strand cross-linking drug, and as such is not wholly unexpected.

When chemical shift data are examined relative to those obtained on assignment of the inter-strand adduct it is noted that chemical shifts at the G-5' end of the intra-strand adduct show a greater similarity to those found in the inter-strand adduct. This provides support for the two-step reaction mechanism postulated by Thurston *et al*,[67] which involved an initial alkylation at the one end of the strand, with a secondary alkylation following in the second reaction step. Presumably the G-5' end of the duplex has reacted first in this case, with the second reaction step producing perturbation on accommodation of the intra-strand cross-link.

The formation of a novel intra-strand cross-link in this manner offers the potential for targeting the human telomere repeat sequence with PBD covalent minor groove binders. Telomeres are the ends of the chromosomes of eukaryotic cells and consist of GT rich repeats in telomere sequence. Normal mortal cells shorten their telomeres during each round of cellular replication, due to an inability of DNA to replicate the last few bases of a sequence. In cancer cells however, the telomerase enzyme is active, and responsible for regenerating the telomere ends, presumably to counteract the rapid cell division that is a feature of cancerous cells. For these reasons telomerase and the telomere repeat sequence have been recognised as an important target for the development of new anti-cancer drugs. Several strategies based on the inhibition of telomerase exist, but it has also been suggested that some classes of DNA damaging drugs might interact preferentially with telomeric sequences. Cis-platin for example, is a G-G cross-linking agent that has been shown to shorten telomeres and inhibit telomerase.[203] The human telomere sequence is -GATTG- and so it is feasible that it could present a potential *binding site for an intra-strand cross-linker such as SJG-136 as shown in* **Figure 142**, with possibilities for alternative length linkers to bind at different sites.

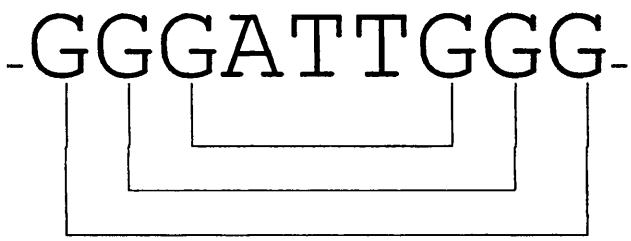


Figure 142: Potential intra-strand cross-link sites on the human telomere repeat sequence

As discussed in **Chapter 1**, previous studies on linker lengths in dimeric PBDs showed the most effective linker units to be those where the number of CH_2 groups (n) was odd (especially 3 or 5).[73,74] Furthermore, the n=5 ligand displayed around 10x greater potency with respect to the n=3 (DSB-120) analogue.[74] The binding of the telomeric repeat sequence is a plausible explanation for the greater potency of the ligand with n=5, as it is possible that this longer sequence, in addition to the inter-strand links reported, could also bind to the telomere repeat sequence *via* an intra-strand cross-link such as that discussed in relation to the SJG-136 adduct. With the considerable current interest in telomeric targeting drugs this possible binding sequence could have major repercussions in anti-cancer drug design.

6.1.3: Future work

Telomere structures are complex and involve the folding of the single stranded DNA into quadruplex formations. It is unlikely that a PBD such as SJG-136 would bind successfully to a quadruplex structure, due to large differences in the shape and size of the grooves. The suggested target site then, is the duplex DNA/RNA hybrid site formed during transcription at the telomerase active site. Binding of PBD drugs to DNA/RNA hybrids has not been assessed, hence future work on PBD adducts will involve the synthesis of DNA/RNA hybrids in order to confirm the possibility of SJG-136 intra-strand adducts with a DNA/RNA hybrid duplex.

6.2: The 5'd(CGATTAATCG)$_2$-adozelesin adduct

Reaction of the 5'*d*(CGATTAATCG)$_2$ duplex with adozelesin has been successfully carried out and has resulted in two adducts in approximately equal proportions. Both adducts were identified and fully assigned using sequential assignment techniques as discussed in **Chapter 2**. One adduct displays a retention of Watson-Crick base pairing throughout the duplex, with minimal disruption of the β-helical structure. The other adduct, however, exhibits Hoogsteen base pairing at the central AT nucleotides, in analogy with previous studies of the symmetrical CPI analogue, bizelesin.[115] Evidence exists for the significant overlap of the benzofuran subunit in both adducts, as shown in **Figure 143** and similarities between the structure of bizelesin compared to overlapped adozelesin molecules have been suggested. There has however, been no evidence to support formation of open base pairing, as seen in the 5'*d*(GTATTAATCG)$_2$-bizelesin adduct.

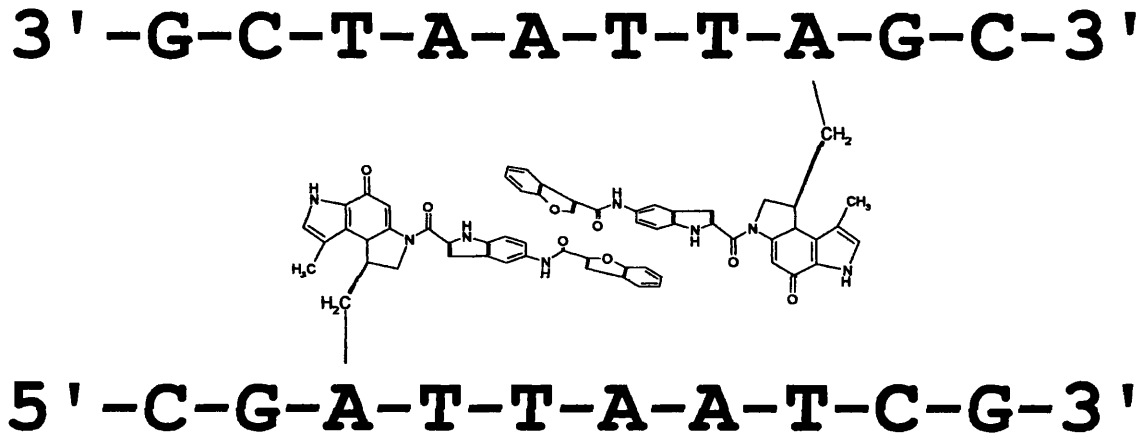


Figure 143: Significant overlap of the benzofuran subunits in an adozelesin 'stacked' *adduct*

The discovery of novel 'stacked' adducts for CPI ligands suggests an alternative model for minor groove drug interactions. Similarities can be drawn between the adozelesin

stacked molecules and analogous stacking in the non-covalent minor groove binding lexitropsin drugs, which have been found to stack in 1:1, 2:1 and 4:1 ratios with the DNA duplex, resulting in a widening of up to 17 Å of the minor groove.[171]

6.2.1: Future work

Possible developments in the area of CPI drug design as a result of discoveries in this project include the synthesis of bulkier DNA interactive ligands. Previous assumptions of a requirement for a planar ligand and a tight fit into the minor groove have been brought into question in light of both these studies, and reports of stacked lexitropsin adducts. It appears that the minor groove is much more accommodating that previously supposed, allowing significant widening to in order to react with bulkier molecules.

Similarities in the shape of the overlapped adozelesin molecules in this study with the shape of the dimeric interstrand cross-linking drug bizelesin also suggest the possibility that stacked adducts may behave as pseudo cross-linkers. This would provide an exciting explanation for the high potency of these compounds compared with other, monomeric alkylating agents.

REFERENCES

1. Mosby's Medical, Nursing and Allied Health Dictionary; 5th ed.; Mosby: Missouri, 1998

2. WHO Melanoma of the skin-statistics; WHO, 2000

3. Foye WO, Sengupta SK (1995) *Principles of Medicinal Chemistry*, 4th ed.; (Williams & Wilkins: USA) pp 826

4. Ammenheuser M, Berenson A, Babiak A, Singleton CR, Whorton EB (1998) *Mutat Res* 403:55-64.

5. Cantor KP, Lynch CF, Hildesheim ME, Dosemeci M, Lubin J (1998) *Epidemiology* 9:21-28.

6. Doyle TJ, Zheng W, Cerhan JR, Hong C-P (1997) *Am J Public Health* 87:1168-1176.

7. Beral V, Herman C, Kay C, Hannaford P, Darby S (1999) *British Medical Journal* 318:96-100.

8. Rushton L (2003) *Occupational and Environmental Medicine* 60:150-156.

9. Levi F (1999) *Eur J Cancer* 35:1046-1058.

10. Little M, Kipriyanov SM, Le Gall F, Moldenhauer G (2000) *Immunol Today* 21:364-370.

11. Fidler IJ (1990) *Cancer Res* 50:6130-6138.

12. Liotta LA, Stracke ML (1988) *Cancer Treatment Res* 40:223-238.

13. Crooke ST (2004) *Annu Rev Med* 55:61-95.

14. Neidle S, Sanderson MR (1983) in *Molecular Aspects of Anti-cancer Drug Action*, eds Neidle S, Waring MJ (Macmillan, London) 1st Ed, pp 35-55.

15. Goodman LS, Wintrobe MM, Dameshek W, Goodman MJ, Gilman A and McLennan MT (1946) *J Am Med Assoc* 105:475-476.

16. Stryer L (1998) *Biochemistry* (WH Freeman, New York) 3rd Ed.

17. Leimgruber W, Stefanovic V, Schenker F, Karr A (1965) *J Am Chem Soc* 87:5791-5793.

18. Leimgruber W, Batcho AD, Schenker F (1965) *J Am Chem Soc* 87:5793-5795.

19. Tendler MD, Korman S (1963) *Nature* 199:501.

20. Grunberg E, Prince HN, Titsworth E, Beskid G, Tendler MD (1966) *Chemotherapy* 11:249-260.

21. Arima K, Kosaka M, Tamura G, Imanaka H, Sakai HJ (1972) *Antibiotics* 25:437-444.

22. Brazhnikova MG, Konstantinova NV, Mesentsev AS (1972) *J Antibiot* 25:668-673.

23. Leber JD, Hoover JRE, Holden KG, Johnson RK, Hecht SMA (1988) *J Am Chem Soc* 110:2992-2993.

24. Takeuchi T, Miyamoto M, Ishizuka M, Naganawa H, Kondo SJ (1976) *Antibiotics* 29:93-96.

25. Kunimoto S, Masuda T, Kanbayashi N, Hamada M, Naganawa HJ (1980) *Antibiotics* 33:665-667.

26. Shimizu K-I, Kawamoto I, Tomita F, Morimoto M, Fujimoto KJ (1982) *Antibiotics* 35:972-978.

27. Kyowa Hakko Kogyo Co Ltd, Japan (1983) Japanese Patent 58180487.

28. Konishi M, Ohkuma H, Naruse N, Kawaguchi HJ (1984) *Antibiotics* 37:200-206.

29. Konishi M, Hatori M, Tomita K, Sugawara M, Ikeda CJ (1984) *Antibiotics* 37:191-199.

30. Hochlowski JE, Andres WW, Theriault RJ, Jackson M, McAlpine JB (1987) *J Antibiot* 40:145-148.

31. Tsunakawa M, Kamei H, Konishi M, Miyaki T, Oki T (1988) *J Antibiot* 41:1366-1373.

32. Hara M, Tamaoki T, Yoshida M, Morimoto M, Nakano H (1988) *J Antibiot* 41:702-704.

33. Kohn KW, Spears CL (1970) *J Mol Biol* 51:551-572.

34. Glaubiger D, Kohn KW, Charney E (1974) *Biochim Biophys Acta* 361:303-311.

35. Kohn KW, Glaubiger D, Spears CL (1974) *Biochim Biophys Acta* 361:288-302.

36. Kaplan DJ, Hurley LH (1981) *Biochemistry* 20:7572-7580.

37. Petrusek RL, Anderson GL, Garner TF, Fannin QL, Kaplan DJ, Zimmer SG, Hurley LH (1981) *Biochemistry* 20:1111-1119.

38. Thurston DE (1993) in *Molecular Aspects of Anti-cancer Drug-DNA Interactions* (Macmillan, Hampshire) 1st Ed, pp 54-88.

39. Baraldi PG, Bovero A, Fruttarolo F, Preti D, Tabrizi MA, Pavani MG, Romagnoli R (2004) *Med Res Rev* 24:475-528.

40. Hurley LH (1977) *J Antibiot* 30:349-370.

41. Kopka ML, Goodsell DS, Baikalov I, Grezeskowiak K, Cascio D, Dickerson RE *(1994) Biochemistry 33:13593-13610.*

42. Rao SN, Singh UC, Kollman PA (1986) *J Med Chem* 29:2484-2492.

43. Remers WA, Mabilia M, Hopfinger AJ (1986) *J Med Chem* 29:2492-2503.

44. Zakrzewska K, Pullman B (1986) *J Biomol Struct Dyn* 4:127-136.

45. Krugh TR, Graves DE, Stone MP (1989) *Biochemistry* 28:9988-9994.

46. Hurley LH, Reck T, Thurston DE, Langley DR, Holden KG, Hertzberg RP, Hoover JRE, Gallagher G, Faucette LF, Mong SM, Johnson RK *(1988) Chem. Res. Toxicol* 1:258-268.

47. Hertzberg RP, Hecht SM, Reynolds VL, Molineux IJ, Hurley LH (1986) *Biochemistry* 25, 1249-1258.

48. Puvvada MS, Hartley JA, Jenkins TC, Thurston DE (1993) *Nucleic Acids Res* 1993, 21, 3671-3675.

49. Pierce JR, Nazimiec M, Tang M (1993) *Biochemistry* 32:7069-7078.

50. Remers WA (1998) in *The Chemistry of Antitumour Antibiotics* (Wiley, New York) Vol 2, 28-92.

51. Horwitz SB, Chang SC, Grollman AP, Borkovec AB (1971) *Science* 147:159-161.

52. Korman S, Tendler MD (1965) *J New Drugs* 5:275-285.

53. Cargill C, Bachmann E, Zbinden G (1974) *J Natl Cancer Inst* 53:481-486.

54. Kaneko T, Wong H, Doyle TW, Rose WC, Bradner WT (1985) *J Med Chem* 28:388-392.

55. Reddy BSP, Sharma SK, Lown JW (2001) *Curr Med Chem* 8:475-508.

56. Jones GB, Davey CL, Jenkins TC, Kamal A, Kneale GG, Neidle S, Webster GD, Thurston DE (1990) *Anti-Cancer Drug Des* 5:249-264.

57. Thurston DE, Bose DS, Howard PW, Jenkins TC, Leoni A, Baraldi PG, Guiotto A, Cacciari B, Kelland LR, Foloppe M-P, Rault S (1999) *J Med Chem* 42:1951-1964.

58. Langlois N, Rojas-Rosseau A, Gaspard C, Werner GH, Darro F, Kiss R (2001) *J Med Chem* 44:3754-3757.

59. Baraldi PG, Leoni A, Cacciari B, Manfredini S, Simoni D, Bergomi M, Menta E, Spinelli S (1994) *J Med Chem* 37:4329-4337.

60. Hopkins PB, Millard TJ, Woo J, Weidner MF, Kirchner JJ, Sigurdson STh, Raucher S (1991) *Tetrahedron* 47:2475-2489.

61. Mattes WB, Hartley JA, Kohn KW, Matheson DW (1988) *Carcinogenesis* 9:2065-2072.

62. Gregson SJ, Howard PW, Hartley JA, Brooks NA, Adams LJ, Jenkins TC, Kelland LR, Thurston DE (2001) *J Med Chem* 44:737-748.

63. Farmer JD, Gustafson GR, Conti A, Zimmt MB, Suggs JW (1991) *Nucleic Acid Res* 19:899-903.

64. Bose DS, Thompson AS, Ching JS, Hartley JA, Berardini MD, Jenkins TC, Neidle S, Hurley LH, Thurston DE (1992) *J Am Chem Soc* 114:4939-4941.

65. Bose DS, Thompson AS, Smellie M, Berardini MD, Hartley JA, Jenkins TC, Neidle S, Thurston DE (1992) *Chem Commun* 1518-1520.

66. Thurston DE, Bose DS, Thompson AS, Howard PW, Leoni A, Croker SJ, Jenkins TC, Neidle S, Hartley JA, Hurley LH (1996) *J Org Chem* 61:8141-8147.

67. Smellie M, Kelland LR, Thurston DE, Souhami RL, Hartley JA (1994) *Br. J. Cancer* 70:48-53.

68. Jenkins TC, Hurley LH, Neidle S, Thurston DE (1994) *J Med Chem* 37:4529-4537.

69. Walton MI, Goddard, P, Kelland LR, Thurston DE, Harrap KR (1996) *Cancer Chemother Pharmacol* 38:431-438.

70. Alley MC, Hollingshead MG, Pacula-Cox CM, Waud WR, Hartley JA, Howard PW, Gregson SJ, Thurston DE, Sausville EA (2004) *Cancer Research* 64:6700-6706

71. Hadjivassileva T, Thurston DE, Taylor PW (2005) *J Antimicrob Chemother* 56:513-518

72. Morris SJ, Thurston DE, Nevell TG (1990) *J Antibiot* 43:1286-1292.

73. Gregson SJ, Howard PW, Gullick DR, Hamaguchi A, Corcoran KE, Brooks NA, Hartley JA, Jenkins TC, Patel S, Guille M, Thurston DE (2004) *J Med Chem* 47:1161-1174.

74. Smellie M, Bose DS, Thompson AS, Jenkins TC, Hartley JA, Thurston DE (2003) *Biochemistry* 42:8232-8239.

75. Hurley L, Draves PH (1993) in *Molecular Aspects of Anticancer Drug-DNA Interactions* (Macmillan, Hampshire) 1st Ed, pp 89–133.

76. Boger DL (1995) *Acc Chem Res* 28:20-29.

77. Aristoff PA (1993) *Adv Med Chem* 2:67-110.

78. Hanka LJ, Dietz A, Gerpheide SA, Kuentzel SL, Martin DG (1978) *J Antibiot* 41:1211-1217.

79. Li LH, Swenson DH, Schpok SL, Kuentzel SL, Dayton BD, Krueger WC (1982) *Cancer Res* 42:999-1004.

80. Martin DG, Biles C, Gerpheide SA, Hanka LJ, Krueger WC, McGovren JP, Miszak SA, Neil GL, Stewart JC, Visser J (1981) *J Antibiot* 34:1119-1125.

81. Hurley LH, Lee C-S, McGovren JP, Warpehoski MA, Mitchell MA, Kelly RC, Aristoff PA (1988) *Biochemistry* 27:3886-3892.

82. Warpehoski MA, Hurley LH (1988) *Chem Res Toxicol* 1:315-333.

83. McGovren JP, Clarke GL, Pratt EA, DeKoning TF (1984) *J Antibiot* 47:64-70.

84. Lee C-S, Sun D, Kizu R, Hurley LH (1991) *Chem Res Toxicol* 4:203-213.

85. Petzold GL, Krueger WC, Swenson DH, Wallace TL, Prairie MD, Li LH (1985) *Proc Am Assoc Cancer Res* 26:225

86. Warpehoski MA (1986) *Tetrahedron Lett* 27:4104-4106.

87. Chidester CG, Krueger WC, Mizsak SA, Duchamp DJ, Martin DG (1981) *J Am Chem Soc* 103:7629-7635.

88. Swenson DH, Li LH, Hurley LH, Rokem JR, Petzold GL, Dayton BD, Wallace TL, Lin AH, Krueger WC (1982) *Cancer Res* 42:2821-2828.

89. Hurley LH, Reynolds VL, Swenson DH, Petzold GL, Scahill TA (1984) *Science* 226:843-844.

90. Scahill TA, Jenson RM, Swenson DH, Hatzenbuhler NT, Petzold G, Wierenga W, Brahme ND (1990) *Biochemistry* 29:2852-2860.

91. Reynolds VL, Molineux IJ, Kaplan DJ, Swenson DH, Hurley LH (1985) *Biochemistry* 24:6228-6247.

92. Boger DL, Invergo BJ, Coleman RS, Zarrinmayeh H, Kitos PA, Thompson SC, Leong T, McLaughlin LW (1990) *Chem-Biol Interact* 74:29-52.

93. Warpehoski MA, Gebhard I, Kelly RC, Krueger WC, Li LH, McGovren JP, Prairie MD, Wicnienski N, Wierenga W (1988) *J Med Chem* 31:590-603.

94. Warpehoski MA, Bradford VS (1988) *Tetrahedron Lett* 29:131-134.

95. Boger DL, Ishizaki T (1990) *Tetrahedron Lett* 31:793-796.

96. Boger DL, Coleman RS, Invergo BJ, Sakya SM, Ishizaki T, Munk SA, Zarrinmayeh H, Kitos PA, Collins Thompson S (1990) *J Am Chem Soc* 112:4623-4632.

97. Boger DL, Munk SA, Ishizaki T (1991) *J Am Chem Soc* 113:2779-2780.

98. Lin CH, Beale JM, Hurley LH (1991) *Biochemistry* 30:3597-3602.

99. McHugh MM, Woynarowski JM, Mitchell MA, Gawron LS, Weiland KL, Beerman TA (1994) *Biochemistry* 33:9158-9168.

100. Zakrewska K, Randrianarivelo M, Pullman B (1987) *Nucleic Acids Res* 15:5775-5785.

101. Lin CH, Hurley LH (1990) *Biochemistry* 29:9503-9507.

102. Lin CH, Patel D (1995) *J Mol Biol* 248:162-179.

103. Skladanowski A, Koba M, Konopa J (2001) *Biochemical Pharmacol* 61:67-72.

104. Cameron L, Thompson AS (2000) *Biochemistry* 39:5004-5012.

105. Liu J-S, Kuo S-R, McHugh MM, Beerman TA, Melendy T (2000) *J Biol Chem* 275:1391-1397.

106. Wold MS (1997) *Annu Rev Biochem* 66:61-92.

107. Bhuyan BK, Smith KS, Adams EG, Wallace TL, Von Hoff DD, Li LH (1992) *Cancer Chemother Pharmacol* 30:348-354.

108. Nguyen HN, Sevin BU, Averette H, Perras J, Ramos R, Donato DJ (1992) *Cancer Res Clin Oncol* 118:515-522.

109. Shamdas GJ, Alberts DS, Modiano M, Wiggins C, Power J, Kasunic DA, Elfring GL, Earhart RH (1994) *Anti Cancer Drugs* 5:10-14.

110. Mitchell MA, Kelly RC, Wicnienski NA, Hatzenbuhler NT, Williams MG, Petzold GC, Slighton JL, Siemieniak PR (1991) *J Am Chem Soc* 113:8994-8995.

111. Ding Z-M, Hurley LH (1991) *Anti-Cancer Drug Des* 6:427-452.

112. Sun D, Hurley LH (1993) *J Am Chem Soc* 115:5925-5933.

113. Lee C-S, Gibson NW (1993) *Biochemistry* 32, 2592-2600.

114. Lee C-S, Gibson NW (1993) *Biochemistry* 32, 9108-9114.

115. Seaman FC, Hurley LH (1993) *Biochemistry* 32:12577-12585.

116. Thompson AS, Sun D, Hurley LH (1995) *J Am Chem Soc* 117:2371-2372.

117. Thompson AS, Hurley LH *J Mol Biol* 1995, 252, 86-101.

118. Wu HM, Crothers DM (1984) *Nature* 308:509-513.

119. Hagerman PJ (1985) *Biochemistry* 24:7033-7037.

120. Koo H-S, Wu HM, Crothers DM (1986) *Nature* 320:501-506.

121. Hagerman PJ (1986) *Nature* 321:449-450.

122. Haran TE, Crothers DE (1989) *Biochemistry* 28:2763-2767.

123. Koo H-S, Crothers DM (1988) *Proc Natl Acad Sci* 85:1763-1767.

124. Nelson HCM, Finch JT, Luisi BF, Klug A (1987) *Nature* 330:221-226.

125. Coll M, Frederick CA, Wang AHJ, Rich A (1987) *Proc Natl Acad Sci* 84:8385-8389.

126. Chuprina VP, Fedoroff OYu, Reid BR (1991) *Biochemistry* 30:561-568.

127. Herrera JE, Chaires JB (1989) *Biochemistry* 28:1993-2000.

128. Chan SS, Breslauer KJ Hogan ME, Kessler DJ, Austin RH, Ojemann J, Passner JM, Wiles NC (1990) *Biochemistry* 29:6161-6171.

129. Thompson AS, Fan J, Sun D, Hansen M, Hurley LH (1995) *Biochemistry* 34:11005-11016.

130. Cao P, McHugh MM, Melendy T, Beerman T (2003) *Mol Cancer Therapeut* 2:651-659.

131. McHugh MM, Kuo SR, Walsh-O'Bierne MH, Liu JS, Melendy T, Beerman TA (1999) *Biochemistry* 38:11508-11515.

132. Woynarowski JM, McHugh MM, Gawron LS, Beerman TA (1995) *Biochemistry* 34:13042-13050.

133. Woynarowski MM (2002) *Biochim Biophys Acta* 1587:300-308.

134. Berezney R, Wei XY (1998) *J Cell Biochem* S30-31:238-242.

135. Tajbakhsh J, Luz H, Bornfleth H, Lampel S, Cremer C, Lichter P (2000) *Exp Cell Res* 255:229-237.

136. Nickerson JA (2001) *J Cell Sci* 114:463-474.

137. Philimonenko VV, Flechon JF, Hozak P (2001) *Exp Cell Res* 264:201-210.

138. Vassetzsky YS, Hair A, Razin SV (2000) *J Cell Biochem* 35:54-60.

139. Kennedy BK, Barbie DA, Classon M, Dyson N, Harlow E (2000) *Genes Dev* 14:2855-2868.

140. Davie JR, Samuel SK, Spencer VA, Holth LT, Chadee DN, Peltier CP, Sun JM, Chon HY, Wright JA (1999) *Biochem Cell Biol* 77:265-275.

141. Woynarowski JM, Beerman TA (1997) *Biochim Biophys Acta* 1353:50-60.

142. Woynarowski JM, Trevino AV, Rodriguez KA, Hardies SC, Benham CJ (2001) *J Biol Chem* 276:40555-40566.

143. Arcamone F, Penco S, Orezzi P, Nicolella V, Pirelli A (1964) *Nature* 1964, 203:1064-1065.

144. Finlay AC, Hochstein FA, Sobin BA, Murphy FX (1951) *J Am Chem Soc* 73:341-343.

145. Hahn FE (1975) in *Antibiotics III* (Corcoran JW, Hahn FE Eds) Springer-Verlag, Berlin-Heidelberg-New York, 79-100.

146. Zimmer CH (1975) *Prog Nucl Acid Res Mol Biol* 15:285-318.

147. Wartell RM, Larson JE, Wells RE (1974) *J Biol Chem* 249:6719-6731.

148. Zimmer C, Wahnert U (1986) *Proc Biophys Mol Biol* 47:31-112.

149. Taylor JS, Schultz PG, Dervan PB (1984) *Tetrahedron* 40:457-465.

150. Dervan PB (1986) *Science* 232:464-471.

151. Klevit RE, Wemmer DE, Reid BR (1986) *Biochemistry* 25:3296-3303.

152. Pelton JG, Wemmer DE (1988) *Biochemistry* 27:8088-8096.

153. Ward B, Rehfuss R, Goodman, J, Dabrowiak JC (1988) *Biochemistry 27:1198-1205.*

154. Fish L, Lane MJ, Vournakis JN (1988) *Biochemistry* 27:6026-6032.

155. Marky LA, Blumenfeld KS, Breslauer KJ (1983) *Nucleic Acids Res* 11:2857-2870.

156. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE (1985) *Proc Natl Acad Sci USA* 82:1376-1380.

157. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE (1985) *J Mol Biol* 183:553-563.

158. Coll M, Ayamani J, van der Marel GA, van Boom JH, Rich A, Wang AJ (1989) *Biochemistry* 28:310-320.

159. Patel DJ (1982) *Proc Natl Acad Sci* 79:6424-6428.

160. Sarma MH, Gupta G, Sarma RH (1985) *J Biomol Struct Dyn* 2:1085-1095.

161. Pardi A, Morden DM, Patel DJ, Tinoco I (1983) *J Am Chem Soc* 22:1107-1113.

162. Neidle S (2001) *Nat Proc Rep* 18:291-309.

163. Zimmer C, Luck G, Thrum H, Pitra C *Eur J Biochem* 26:81-89.

164. Zakrewska K, Lavery R, Pullman B (1983) *Nucleic Acids Res* 11:8825-8839.

165. Breslauer KJ, Remeta DP, Chou W-Y, Ferrante R, Curry J, Zaunczkowski D, Snyder JG, Marky LA (1987) *Proc Natl Acad Sci* 84:8922-8926.

166. Reinert KE, Thrum H (1970) *Stud Biophys* 24/25:319-325.

167. Van Dyke MW, Hertzberg RP, Dervan PB (1982) *Proc Natl Acad Sci* 79:5470-5474.

168. Lown JW (1994) *J Mol Recognit* 7:79-88.

169. Lown JW, Krowicki K, Bhat UG, Skorobogaty A, Ward B, Dabrowiak JC (1986) *Biochemistry* 25:7408-7416.

170. Pelton JG, Wemmer DE (1989) *Proc Natl Acad Sci* 86:5723-5727.

171. Blasko A, Bruice TC (1993) *Proc Natl Acad Sci* 90:10018-10022.

172. Ceccarelli C, Jeffrey GA, Taylor R (1981) *J Mol Struct* 70:255-271.

173. Jeffrey GA, Mitra J (1984) *J Am Chem Soc* 106:5546-5553.

174. Lavery R, Pullman A, Pullman B (1982) *Theor Chim Acta* 62:93-106.

175. Kamal A, Reddy DR, Murali Mohan Reddy PS, Rajendar (2006) *Bioorganic & Med Chem Lett* 16:1160-1163

176. Kamal A, Babu AH, Ramana AV, Ramana KV, Bharathi EV, Kumar MS (2005) *Bioorganic & Med. Chem Lett* 15:2621-2623

177. Wang J-J, Shen Y-K, Hu W-P, Hsieh M-C, Lin F-L, Hsu M-K, Hsu M-H, (2006) *J Med Chem* 49:1442-1449

178. Wells G, Martin CRH, Howard PW, Sands ZA, Laughton CA, Tiberghien A, Woo CK, Masterson LA, Stephenson MJ, Hartley JA, Jenkins TC, Shnyder SD, Loadman PM, Waring MJ, Thurston DE (2006) *J Med Chem* 49:5442-5461

179. Cameron L (1999) in *Adozelesin – DNA Binding by ^{1}H NMR.* PhD Thesis, University of Bath.

180. Cameron L, Thompson AS Unpublished data, personal communication, University of Bath

181. Mountzouris JA, Wang J, Thurston D, Hurley LH (1994) *J Med Chem* 37:3132-3140.

182. *Applied Biosystems User's Manual – DNA Synthesizer Model 381A* (1987) Version 1.23.

183. Primrose WU (1993) in *NMR of Macromolecules A Practical Approach*, ed Roberts GCK (IRL Press, Oxford) pp 7-34.

184. Sanders JKM, Hunter BK (1993) in *Modern NMR Spectroscopy – a guide for chemists* (Oxford University Press, Oxford) 2nd Ed, pp 11-24 & 164-166.

185. Hore PJ (2004) *Nuclear Magnetic Resonance* (Oxford University Press).

186. Kemp W (1991) in *Organic Spectroscopy* (Palgrave Macmillan, Hampshire) 3rd Ed, pp 152-154.

187. Karplus MJ (1961) *Phys Chem* 64:1793-1798.

188. Figure taken from www.chemistry.ccsu.edu/glagovich/teaching/316/nmr/couplingbasics.html

189. Wuthrich K (1986) *NMR of Proteins and Nucleic Acids,* (Wiley, New York)

190. Wijmenga SS, Mooren MMW, Hilbers CW (1993) in *NMR of Macromolecules – a Practical Approach,* ed Roberts GCK (IRL Press, Oxford) pp 217-288.

191. Hare DR, Wemmer DE, Chou S-H, Drobny G (1983) *J Mol Biol* 171:319-336.

192. Chary KVR, Hosur RV, Govil G, Zu-kun T, Miles HT (1987) *Biochemistry* 26:1315-1322.

193. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) *J Comput Chem* 25:1605-1612

194. Barkley MD, Cheatham S, Thurston DE, Hurley LH (1986) *Biochemistry* 25:3021-3031

195. Graves DE, Pattaroni C, Krishnan BS, Ostrander JM, Hurley LH, Krugh TR (1984) *J Biol Chem* 259:8202-8209

196. Reid BR (1987) *Q Rev Biophys* 20:1-33

197. SYBYL® Version 6.2 Manual (1995) Tripos Inc., St. Louis, Missouri, USA

198. HyperChem(TM), Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA

199. Westhof E, Rubin-Carrex C, Fritsch V (1995) in *Computer Modeling in Molecular Biology* (Goodfellow JM ed) VCH Verlagsgesellschaft, Weinheim (Federal Republic of Germany) and VCH publishers, New York NY (USA) pp1-7

200. Walker GS, Fagerness PE, Farley KA, Misak SA (1997) *J Heterocycl Chem* 34:295

201. Wang AH-J, Ughetto G, Quigley G-J, Hakoshima T, van der Marel G, van Boom JH, Rich A (1984) *Science* 225:1115-1121

202. Wang A H-J, Ughetto G, Quigley G-J, Rich A (1986) *J Biomol Struct Dyn* 4:319-342

203. Zhang RG, Zhang RP, Wang XW, Xie H (2002) *Cell Res* 12:55-62

Appendix I

Master Table Showing DNA-DNA Connectivities and Restraints for the 5'*d*-(CICGATCICG)$_2$-SJG-136 Adduct

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper Limit	Lower Limit
1	I2-H8	C3-H6	8.28	7.34	W	4.50	3.50
2	C3-H6	G4-H8	7.34	7.80	W	4.50	3.50
3	G4-H8	A5-H8	7.80	8.15	W	4.50	3.50
4	A5-H8	T6-H6	8.15	7.05	W	4.50	3.50
5	C7-H6	I8-H8	7.14	8.12	W	4.50	3.50
6	I8-H8	C9-H8	8.12	7.20	W	4.50	3.50
7	C1-H6	C1-H1'	7.90	5.52	M	3.50	2.00
8	I2-H8	I2-H1'	8.28	6.15	M	3.50	2.00
9	C3-H6	C3-H1'	7.34	5.47	M	3.50	2.00
10	G4-H8	G4-H1'	7.80	6.08	M	3.50	2.00
11	A5-H8	A5-H1'	8.15	6.30	W	4.50	3.50
12	T6-H6	T6-H1'	7.05	5.71	W/M	4.00	3.00
13	C7-H6	C7-H1'	7.14	5.44	W	4.50	3.50
14	I8-H8	I8-H1'	8.12	5.84	M	3.50	2.00
15	C9-H8	C9-H1'	7.20	5.77	M	3.50	2.00
16	G10-H8	G10-H1'	7.90	6.06	M	3.50	2.00
17	C1-H1'	I2-H8	5.52	8.28	W	4.50	3.50
18	I2-H1'	C3-H6	6.15	7.34	M	3.50	2.00
19	C3-H1'	G4-H8	5.47	7.80	W/M	4.00	3.00
20	G4-H1'	A5-H8	6.08	8.15	M	3.50	2.00
21	A5-H1'	T6-H6	6.30	7.05	M	3.50	2.00
22	T6-H1'	C7-H6	5.71	7.14	M/W	4.00	3.00
23	C7-H1'	I8-H8	5.44	8.12	W	4.50	3.50
24	I8-H1'	C9-H8	5.84	7.20	M	3.50	2.00
25	C9-H1'	G10-H8	5.77	7.90	W	4.50	3.50
26	C1-H6	C1-H2'	7.60	2.28	M	3.50	2.00
27	C1-H6	C1-H2''	7.60	1.82	S	1.00	2.00
28	C1-H2'	I2-H8	2.28	8.28	M	3.50	2.00
29	C1-H2''	I2-H8	1.82	8.28	M	3.50	2.00
30	I2-H8	I2-H2'	8.28	2.82	S	1.00	2.00
31	I2-H8	I2-H2''	8.28	2.66	S	1.00	2.00
32	I2-H2'	C3-H6	2.82	7.34	M	3.50	2.00
33	I2-H2''	C3-H6	2.66	7.34	M	3.50	2.00
34	C3-H6	C3-H2'	7.34	2.27	M	3.50	2.00
35	C3-H6	C3-H2''	7.34	1.86	S	1.00	2.00
36	C3-H2'	G4-H8	2.27	7.80	M	3.50	2.00
37	C3-H2''	G4-H8	1.86	7.80	M	3.50	2.00
38	G4-H8	G4-H2'	7.80	2.94	M	3.50	2.00
39	G4-H8	G4-H2''	7.80	2.66	S	1.00	2.00
40	G4-H2'	A5-H8	2.94	8.15	S	1.00	2.00
41	G4-H2''	A5-H8	2.66	8.15	S	1.00	2.00

42	A5-H8	A5-H2'	8.15	2.66	S	1.00	2.00
43	A5-H8	A5-H2''	8.15	2.46	S	1.00	2.00
44	A5-H2'	T6-H6	2.66	7.05	M	3.50	2.00
45	A5-H2''	T6-H6	2.46	7.05	M	3.50	2.00
46	T6-H6	T6-H2'	7.05	2.17	M	3.50	2.00
47	T6-H6	T6-H2''	7.05	1.94	S	1.00	2.00
48	T6-H2'	C7-H6	2.17	7.14	M	3.50	2.00
49	T6-H2''	C7-H6	1.94	7.14	M	3.50	2.00
50	C7-H6	C7-H2'	7.14	2.09	M	3.50	2.00
51	C7-H6	C7-H2''	7.14	1.65	M	3.50	2.00
52	C7-H2'	I8-H8	2.09	8.12	M	3.50	2.00
53	C7-H2''	I8-H8	1.65	8.12	M/W	4.00	3.00
54	I8-H8	I8-H2''	8.12	2.10	M	3.50	2.00
55	I8-H2'	C9-H6	2.57	7.20	M	3.50	2.00
56	I8-H2''	C9-H6	2.10	7.20	M	3.50	2.00
57	C9-H6	C9-H2'	7.20	2.13	M	3.50	2.00
58	C9-H6	C9-H2''	7.20	1.46	S	1.00	2.00
59	C9-H2'	G10-H8	2.13	7.90	M	3.50	2.00
60	C9-H2''	G10-H8	1.46	7.90	M	3.50	2.00
61	G10-H8	G10-H2'	7.90	2.55	S	1.00	2.00
62	G10-H8	G10-H2''	7.90	2.26	S	1.00	2.00
63	C1-H6	C1-H3'	7.60	4.61	W	4.50	3.50
64	I2-H8	I2-H3'	8.28	4.91	M	3.50	2.00
65	C3-H6	C3-H3'	7.34	4.92	VW	5.00	4.00
66	G4-H8	G4-H3'	7.80	4.95	M	3.50	2.00
67	A5-H8	A5-H3'	8.15	4.88	M/W	4.00	3.00
68	T6-H6	T6-H3'	7.05	4.55	M	3.50	2.00
69	I8-H8	I8-H3'	8.12	4.83	W	4.50	3.50
70	G10-H8	G10-H3'	7.90	4.58	M	3.50	2.00
71	C3-H3'	G4-H8	4.92	7.80	W	4.50	3.50
72	G4-H3'	A5-H8	4.95	8.15	M/W	4.00	3.00
73	A5-H3'	T6-H6	4.88	7.05	VW	5.00	4.00
74	C1-H6	C1-H4'	7.60	3.63	M	3.50	2.00
75	I2-H8	I2-H4'	8.28	4.32	W	4.50	3.50
76	C3-H6	C3-H4'	7.34	4.06	M	3.50	2.00
77	G4-H8	G4-H4'	7.80	4.06	M	3.50	2.00
78	A5-H8	A5-H4'	8.15	3.94	W	4.50	3.50
79	T6-H6	T6-H4'	7.05	3.98	M	3.50	2.00
80	C7-H6	C7-H4'	7.14	4.02	W	4.50	3.50
81	C9-H8	C9-H4'	7.20	3.94	W	4.50	3.50
82	G10-H8	G10-H4'	7.90	3.96	M	3.50	2.00
83	C1-H1'	C1-H3'	5.52	8.28	W	4.50	3.50
84	I2-H1'	I2-H3'	6.15	7.34	M	3.50	2.00
85	G4-H1'	G4-H3'	6.08	8.15	M	3.50	2.00
86	A5-H1'	A5-H3'	6.30	7.05	M	3.50	2.00
87	I8-H1'	I8-H3'	5.84	7.20	M	3.50	2.00
88	G10-H1'	G10-H3'	6.06	4.58	M	3.50	2.00
89	C1-H1'	C1-H4'	5.52	3.63	W	4.50	3.50

90	I2-H1'	I2-H4'	6.15	4.32	M	3.50	2.00
91	C3-H1'	C3-H4'	5.47	4.06	M	3.50	2.00
92	G4-H1'	G4-H4'	6.08	4.06	S	1.00	2.00
93	A5-H1'	A5-H4'	6.30	3.94	M	3.50	2.00
94	T6-H1'	T6-H4'	5.71	3.98	M	3.50	2.00
95	C7-H1'	C7-H4'	5.44	4.02	M	3.50	2.00
96	I8-H1'	I8-H4'	5.84	3.82	M/S	3.00	1.00
97	C9-H1'	C9-H4'	5.77	3.94	M	3.50	2.00
98	G10-H1'	G10-H4'	6.06	3.96	M	3.50	2.00
99	C1-H4'	C1-H2'	3.63	2.28	W	4.50	3.50
100	I2-H4'	I2-H2'	4.32	2.82	W	4.50	3.50
101	C3-H4'	C3-H2'	4.06	2.27	M	3.50	2.00
102	G4-H4'	G4-H2'	4.06	2.94	M	3.50	2.00
103	A5-H4'	A5-H2'	3.94	2.66	M	3.50	2.00
104	T6-H4'	T6-H2'	3.98	2.17	M	3.50	2.00
105	C7-H4'	C7-H2'	4.02	2.09	W/M	4.00	3.00
106	I8-H4'	I8-H2'	3.82	2.57	M	3.50	2.00
107	C9-H4'	C9-H2'	3.94	2.13	M	3.50	2.00
108	G10-H4'	G10-H2'	3.96	2.55	M	3.50	2.00
109	C1-H4'	C1-H2''	3.63	1.82	M	3.50	2.00
110	I2-H4'	I2-H2''	4.32	2.66	W	4.50	3.50
111	C3-H4'	C3-H2''	4.06	1.86	M	3.50	2.00
112	G4-H4'	G4-H2''	4.06	2.66	M	3.50	2.00
113	A5-H4'	A5-H2''	3.94	2.46	M	3.50	2.00
114	T6-H4'	T6-H2''	3.98	1.94	M	3.50	2.00
115	C7-H4'	C7-H2''	4.02	1.65	W/M	4.00	3.00
116	I8-H4'	I8-H2''	3.82	2.10	M	3.50	2.00
117	C9-H4'	C9-H2''	3.94	1.46	M	3.50	2.00
118	G10-H4'	G10-H2''	3.96	2.26	M	3.50	2.00
119	C1-H3'	C1-H4'	5.52	3.63	W	4.50	3.50
120	I2-H3'	I2-H4'	6.15	4.32	M	3.50	2.00
121	C3-H3'	C3-H4'	5.47	4.06	M	3.50	2.00
122	G4-H3'	G4-H4'	6.08	4.06	M	3.50	2.00
123	A5-H3'	A5-H4'	6.30	3.94	M	3.50	2.00
124	I8-H3'	I8-H4'	5.84	3.82	M	3.50	2.00
125	G10-H3'	G10-H4'	6.06	3.96	W/M	4.00	3.00
126	T6-CH3	A5-H8	1.20	8.15	S	1.00	2.00
127	T6-CH3	T6-H6	1.20	7.05	S	1.00	2.00
128	T6-CH3	A5-H1'	1.20	6.30	M	3.50	2.00
129	T6-CH3	C7-H5	1.20	5.26	W	4.50	3.50
130	T6-CH3	A5-H3'	1.20	4.88	M/W	4.00	3.00
131	T6-CH3	A5-H2'	1.20	2.66	S	1.00	2.00
132	T6-CH3	A5-H2''	1.20	2.46	S	1.00	2.00
133	I2-H2	I8-H2	7.82	7.38	W	4.50	3.50
134	I2-H2	G10-H1'	7.82	6.06	M	3.50	2.00
135	I2-H2	C9-H1'	7.82	5.77	W	4.50	3.50
136	I2-H2	I2-H1'	7.82	6.15	M	3.50	2.00
137	I8-H2	I8-H1'	7.38	5.84	M	3.50	2.00

138	I8-H2	C9-H4'	7.38	3.94	M	3.50	2.00
139	I2-H2	C3-H1'	7.82	5.47	M	3.50	2.00
140	I8-H2	C9-H1'	7.38	5.77	M	3.50	2.00
141	A5-H2	T6-H1'	8.11	5.71	W	4.50	3.50
142	A5-H2	C7-H1'	8.11	5.44	W	4.50	3.50
143	C1-H5	C1-H4'	5.80	3.63	VW	5.00	4.00
144	C1-H5	C1-H2'	5.80	2.28	W	4.50	3.50
145	C1-H5	C1-H2''	5.80	1.82	M	3.50	2.00
146	C3-H5	I2-H2'	5.29	2.82	M	3.50	2.00
147	C3-H5	I2-H2''	5.29	2.66	M	3.50	2.00
148	C3-H5	C3-H2'	5.29	2.27	M	3.50	2.00
149	C3-H5	C3-H2''	5.29	1.86	M	3.50	2.00
150	C7-H5	C7-H2'	5.27	2.09	M	3.50	2.00
151	C7-H5	C7-H2''	5.27	1.65	M	3.50	2.00
152	C9-H5	I8-H2'	5.26	2.57	M	3.50	2.00
153	C9-H5	I8-H2''	5.26	2.10	M	3.50	2.00
154	C9-H5	C9-H2'	5.26	2.13	M	3.50	2.00
155	C9-H5	C9-H2''	5.26	1.46	M	3.50	2.00

Master Table Showing Drug-DNA Connections for the 5'*d*-(CICGATCICG)$_2$- SJG-136 Adduct

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper Limit	Lower Limit
1	H1a	I8-H3'	2.55	4.83	M	3.50	2.00
2	H1a	I8-H2	2.55	7.38	M	3.50	2.00
3	H1b	I8-H3'	3.21	4.83	M/W	4.00	3.00
4	H1b	C7-H1'	3.21	5.44	W	4.50	3.50
5	H1b	I8-H2	3.21	7.38	M	3.50	2.00
6	H2a/b	A5-H4'	5.84	3.94	M	3.50	2.00
7	H2a/b	C9-H4'	5.84	3.94	M	3.50	2.00
8	H2a/b	G4-H4'	5.84	4.06	W	4.50	3.50
9	H2a/b	I8-H2	5.84	7.38	W	4.50	3.50
10	H3a/b	I8-H1'	4.02/4.06	5.84	W	4.50	3.50
11	H3a/b	A5-H3'	4.02/4.06	4.88	M	3.50	2.00
12	H3a/b	A5-H1'	4.02/4.06	6.30	W	4.50	3.50
13	H3a/b	C9-H1'	4.02/4.06	5.77	W	4.50	3.50
14	H6	T6-H4'	7.05	3.98	W/M	4.00	3.00
15	H9	C7-H1'	7.05	5.44	W	4.50	3.50
16	H9	C7-H4'	6.53	4.02	W	4.50	3.50
17	H9	A5-H2	6.53	8.11	S	2.00	1.00
18	H9	T6-H1'	6.53	5.71	M	3.50	2.00
19	H11	I8-H2	5.64	7.38	W	4.50	3.50
20	H11	G4-H1'	5.64	6.28	M	3.50	2.00
21	H11	A5-H1'	5.64	6.30	M	3.50	2.00
22	H11a	I8-H2	3.92	7.38	M	3.50	2.00

23	H11a	I8-H1'	3.92	5.84	M	3.50	2.00
24	H11a	C7-H1'	3.92	5.44	W	4.50	3.50
25	H12a/b	C7-H1'	4.24	5.44	M/W	4.00	3.00
26	H12a/b	T6-H1'	4.24	5.71	M	3.50	2.00
27	H12a/b	A5-H2	4.24	8.11	M	3.50	2.00
28	H13a/b	C7-H1'	2.28	5.44	M	3.50	2.00
29	H13a/b	C7-H4'	2.28	4.02	S	2.00	1.00
30	H13a/b	T6-H1'	2.28	5.71	M	3.50	2.00
31	H14	C7-H1'	3.83	5.44	W	4.50	3.50

Appendix II

Master Table Showing All DNA-DNA Connections for the 5'*d*-CTCATCAC.GTGATGAG-SJG-136 Intra-strand Adduct.

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper Limit	Lower Limit
1	C1-H6	T2-H6	7.76	7.52	M/S	3.00	1.00
2	T2-H6	C3-H6	7.52	7.45	M/S	3.00	1.00
3	C3-H6	A4-H8	7.45	8.32	M	3.50	2.00
4	A4-H8	T5-H6	8.32	6.89	W	4.50	3.50
5	T5-H6	C6-H6	6.89	7.07	M	3.50	2.00
6	C6-H6	A7-H8	7.07	8.04	M	3.50	2.00
7	A7-H8	C8-H6	8.04	7.30	M	3.50	2.00
8	G9-H8	T10-H6	7.83	7.30	W	4.50	3.50
9	T10-H6	G11-H8	7.30	7.78	W/M	4.00	3.00
10	G11-H8	A12-H8	7.78	8.14	M	3.50	2.00
11	A12-H8	T13-H6	8.14	7.01	W	4.50	3.50
12	T13-H6	G14-H8	7.01	7.52	M	3.50	2.00
13	G14-H8	A15-H8	7.52	7.83	M	3.50	2.00
14	A15-H8	G16-H8	7.83	7.51	M	3.50	2.00
15	C1-H6	C1-H1'	7.76	5.80	S	2.00	1.00
16	T2-H6	T2-H1'	7.52	5.93	M	3.50	2.00
17	C3-H6	C3-H1'	7.45	5.99	M	3.50	2.00
18	A4-H8	A4-H1'	8.32	6.29	W	4.50	3.50
19	T5-H6	T5-H1'	6.89	5.55	M/W	4.00	3.00
20	C6-H6	C6-H1'	7.07	5.61	M	3.50	2.00
21	A7-H8	A7-H1'	8.04	5.88	M	3.50	2.00
22	C8-H6	C8-H1'	7.30	5.97	M	3.50	2.00
23	G9-H8	G9-H1'	7.83	5.94	W	4.50	3.50
24	T10-H6	T10-H1'	7.30	5.44	M	3.50	2.00
25	G11-H8	G11-H1'	7.78	6.08	W/M	4.00	3.00
26	A12-H8	A12-H1'	8.14	6.24	W	4.50	3.50
27	T13-H6	T13-H1'	7.01	5.40	M	3.50	2.00
28	G14-H8	G14-H1'	7.52	5.94	W	4.50	3.50
29	A15-H8	A15-H1'	7.83	5.77	W	4.50	3.50
30	G16-H8	G16-H1'	7.51	5.92	M	3.50	2.00
31	T2-H1'	C1-H6	7.52	7.76	M	3.50	2.00
32	C3-H1'	T2-H6	7.45	7.52	M	3.50	2.00
33	A4-H1'	C3-H6	8.32	7.45	M	3.50	2.00
34	T5-H1'	A4-H8	6.89	8.32	M	3.50	2.00
35	C6-H1'	T5-H6	7.07	6.89	M	3.50	2.00
36	A7-H1'	C6-H6	8.04	7.07	M	3.50	2.00
37	C8-H1'	A7-H8	7.30	8.04	M	3.50	2.00
38	T10-H1'	G9-H8	7.83	7.83	M	3.50	2.00
39	G11-H1'	T10-H6	7.30	7.30	M	3.50	2.00
40	A12-H1'	G11-H8	7.78	7.78	M	3.50	2.00

41	T13-H1'	A12-H8	8.14	8.14	M	3.50	2.00
42	G14-H1'	T13-H6	7.01	7.01	M/S	3.00	1.00
43	A15-H1'	G14-H8	7.52	7.52	M	3.50	2.00
44	G16-H1'	A15-H8	7.83	7.83	M	3.50	2.00
45	C1-H6	C1-H2'	7.51	2.44	M/S	3.00	1.00
46	T2-H6	T2-H2'	7.52	2.44	S	2.00	1.00
47	C3-H6	C3-H2'	7.45	2.55	S	2.00	1.00
48	A4-H8	A4-H2'	8.32	2.75	S	2.00	1.00
49	T5-H6	T5-H2'	6.89	2.18	S	2.00	1.00
50	C6-H6	C6-H2'	7.07	2.12	S	2.00	1.00
51	A7-H8	A7-H2'	8.04	2.53	M	3.50	2.00
52	C8-H6	C8-H2'	7.30	2.02	S	2.00	1.00
53	G9-H8	G9-H2'	7.83	2.72	S	2.00	1.00
54	T10-H6	T10-H2'	7.30	2.42	S	2.00	1.00
55	G11-H8	G11-H2'	7.78	2.84	S	2.00	1.00
56	A12-H8	A12-H2'	8.14	2.55	S	2.00	1.00
57	T13-H6	T13-H2'	7.01	2.15	S	2.00	1.00
58	G14-H8	G14-H2'	7.52	2.57	S	2.00	1.00
59	A15-H8	A15-H2'	7.83	2.55	S	2.00	1.00
60	G16-H8	G16-H2'	7.51	2.25	S	2.00	1.00
61	C1-H6	C1-H2''	7.76	2.08	M	3.50	2.00
62	T2-H6	T2-H2''	7.52	2.15	S	2.00	1.00
63	C3-H6	C3-H2''	7.45	2.12	S	2.00	1.00
64	A4-H8	A4-H2''	8.32	2.60	M	3.50	2.00
65	T5-H6	T5-H2''	6.89	1.76	M	3.50	2.00
66	C6-H6	C6-H2''	7.07	1.64	S	2.00	1.00
67	A7-H8	A7-H2''	8.04	2.42	S	2.00	1.00
68	C8-H6	C8-H2''	7.30	1.97	M	3.50	2.00
69	G9-H8	G9-H2''	7.83	2.47	M	3.50	2.00
70	T10-H6	T10-H2''	7.30	2.22	S	2.00	1.00
71	G11-H8	G11-H2''	7.78	2.60	M/S	3.00	1.00
72	A12-H8	A12-H2''	8.14	2.44	S	2.00	1.00
73	T13-H6	T13-H2''	7.01	2.10	S	2.00	1.00
74	G14-H8	G14-H2''	7.52	2.45	S	2.00	1.00
75	A15-H8	A15-H2''	7.83	2.44	S	2.00	1.00
76	G16-H8	G16-H2''	7.51	2.08	S	2.00	1.00
77	T2-H6	C1-H2'	7.52	2.44	S	2.00	1.00
78	C3-H6	T2-H2'	7.45	2.44	S	2.00	1.00
79	A4-H8	C3-H2'	8.32	2.55	M	3.50	2.00
80	T5-H6	A4-H2'	6.89	2.75	M	3.50	2.00
81	C6-H6	T5-H2'	7.07	2.18	M	3.50	2.00
82	A7-H8	C6-H2'	8.04	2.12	M	3.50	2.00
83	C8-H6	A7-H2'	7.30	2.53	S	2.00	1.00
84	T10-H6	G9-H2'	7.30	2.72	M	3.50	2.00
85	G11-H8	T10-H2'	7.78	2.42	M	3.50	2.00
86	A12-H8	G11-H2'	8.14	2.84	S	2.00	1.00
87	T13-H6	A12-H2'	7.01	2.55	M/S	3.00	1.00
88	G14-H8	T13-H2'	7.52	2.15	S	2.00	1.00

89	A15-H8	G14-H2'	7.83	2.57	M/S	3.00	1.00
90	G16-H8	A15-H2'	7.51	2.55	S	2.00	1.00
91	T2-H6	C1-H2''	7.52	2.08	S	2.00	1.00
92	C3-H6	T2-H2''	7.45	2.15	M	3.50	2.00
93	A4-H8	C3-H2''	8.32	2.12	S	2.00	1.00
94	T5-H6	A4-H2''	6.89	2.60	M/S	3.00	1.00
95	C6-H6	T5-H2''	7.07	1.76	S	2.00	1.00
96	A7-H8	C6-H2''	8.04	1.64	M/S	3.00	1.00
97	C8-H6	A7-H2''	7.30	2.42	M/S	3.00	1.00
98	T10-H6	G9-H2''	7.30	2.47	S	2.00	1.00
99	G11-H8	T10-H2''	7.78	2.22	S	2.00	1.00
100	A12-H8	G11-H2''	8.14	2.60	S	2.00	1.00
101	T13-H6	A12-H2''	7.01	2.44	S	2.00	1.00
102	G14-H8	T13-H2''	7.52	2.10	S	2.00	1.00
103	A15-H8	G14-H2''	7.83	2.45	S	2.00	1.00
104	G16-H8	A15-H2''	7.51	2.44	S	2.00	1.00
105	C1-H6	C1-H3'	7.76	4.55	M	3.50	2.00
106	T2-H6	T2-H3'	7.52	4.79	W	4.50	3.50
107	C3-H6	C3-H3'	7.45	4.82	W	4.50	3.50
108	A4-H8	A4-H3'	8.32	4.89	W	4.50	3.50
109	T5-H6	T5-H3'	6.89	4.51	M	3.50	2.00
110	A7-H8	A7-H3'	8.04	4.78	W	4.50	3.50
111	G9-H8	G9-H3'	7.83	4.71	W	4.50	3.50
112	T10-H6	T10-H3'	7.30	4.82	W	4.50	3.50
113	G11-H8	G11-H3'	7.78	4.97	M	3.50	2.00
114	A12-H8	A12-H3'	8.14	4.86	W	4.50	3.50
115	T13-H6	T13-H3'	7.01	4.60	W	4.50	3.50
116	G14-H8	G14-H3'	7.52	4.55	M/W	4.00	3.00
117	A15-H8	A15-H3'	7.83	4.87	W	4.50	3.50
118	G16-H8	G16-H3'	7.51	4.47	M	3.50	2.00
119	T2-H6	C1-H3'	7.52	4.55	M	3.50	2.00
120	C3-H6	T2-H3'	7.45	4.79	W	4.50	3.50
121	A4-H8	C3-H3'	8.32	4.82	W	4.50	3.50
122	T5-H6	A4-H3'	6.89	4.89	W	4.50	3.50
123	C6-H6	T5-H3'	7.07	4.51	W	4.50	3.50
124	G11-H8	T10-H3'	7.78	4.82	VW	5.00	4.00
125	A12-H8	G11-H3'	8.14	4.97	W	4.50	3.50
126	T13-H6	A12-H3'	7.01	4.86	W	4.50	3.50
127	G14-H8	T13-H3'	7.52	4.60	VW	5.00	4.00
128	A15-H8	G14-H3'	7.83	4.55	W	4.50	3.50
129	G16-H8	A15-H3'	7.51	4.87	VW	5.00	4.00
130	C1-H6	C1-H4'	7.76	3.98	M/W	4.00	3.00
131	T2-H6	T2-H4'	7.52	4.13	M/W	4.00	3.00
132	C3-H6	C3-H4'	7.45	4.18	M/W	4.00	3.00
133	A4-H8	A4-H4'	8.32	3.64	W	4.50	3.50
134	T5-H6	T5-H4'	6.89	3.56	VW	5.00	4.00
135	A7-H8	A7-H4'	8.04	3.62	VW	5.00	4.00
136	C8-H6	C8-H4'	7.30	4.04	M/W	4.00	3.00

137	G9-H8	G9-H4'	7.83	4.13	W	4.50	3.50
138	T10-H6	T10-H4'	7.30	4.13	M/W	4.00	3.00
139	G11-H8	G11-H4'	7.78	4.35	W	4.50	3.50
140	A12-H8	A12-H4'	8.14	3.86	W	4.50	3.50
141	T13-H6	T13-H4'	7.01	3.30	W	4.50	3.50
142	G14-H8	G14-H4'	7.52	4.05	M	3.50	2.00
143	A15-H8	A15-H4'	7.83	4.05	M/W	4.00	3.00
144	G16-H8	G16-H4'	7.51	3.98	M	3.50	2.00
145	C1-H1'	C1-H3'	5.80	4.55	M	3.50	2.00
146	T2-H1'	T2-H3'	5.93	4.79	M	3.50	2.00
147	C3-H1'	C3-H3'	5.99	4.82	M	3.50	2.00
148	A4-H1'	A4-H3'	6.29	4.89	M	3.50	2.00
149	T5-H1'	T5-H3'	5.55	4.51	M	3.50	2.00
150	A7-H1'	A7-H3'	5.88	4.78	M	3.50	2.00
151	G9-H1'	G9-H3'	5.94	4.71	W	4.50	3.50
152	T10-H1'	T10-H3'	5.44	4.82	M/W	4.00	3.00
153	G11-H1'	G11-H3'	6.08	4.97	M	3.50	2.00
154	A12-H1'	A12-H3'	6.24	4.86	M	3.50	2.00
155	T13-H1'	T13-H3'	5.40	4.60	W	4.50	3.50
156	G14-H1'	G14-H3'	5.94	4.55	M	3.50	2.00
157	A15-H1'	A15-H3'	5.77	4.87	M	3.50	2.00
158	G16-H1'	G16-H3'	5.92	4.47	M	3.50	2.00
159	C1-H1'	C1-H4'	5.80	3.98	M	3.50	2.00
160	T2-H1'	T2-H4'	5.93	4.13	M	3.50	2.00
161	C3-H1'	C3-H4'	5.99	4.18	M	3.50	2.00
162	A4-H1'	A4-H4'	6.29	3.64	M	3.50	2.00
163	T5-H1'	T5-H4'	5.55	3.56	M	3.50	2.00
164	C6-H1'	C6-H4'	5.61	3.04	M	3.50	2.00
165	A7-H1'	A7-H4'	5.88	3.62	M/S	3.00	1.00
166	C8-H1'	C8-H4'	5.97	4.04	M	3.50	2.00
167	G9-H1'	G9-H4'	5.94	4.13	S	2.00	1.00
168	T10-H1'	T10-H4'	5.44	4.13	M	3.50	2.00
169	G11-H1'	G11-H4'	6.08	4.35	M/S	3.00	1.00
170	A12-H1'	A12-H4'	6.24	3.86	M/S	3.00	1.00
171	T13-H1'	T13-H4'	5.40	3.30	VW	5.00	4.00
172	G14-H1'	G14-H4'	5.94	4.05	M	3.50	2.00
173	A15-H1'	A15-H4'	5.77	4.05	M	3.50	2.00
174	G16-H1'	G16-H4'	5.92	3.98	S	2.00	1.00
175	C1-H2'	C1-H4'	2.44	3.98	S	2.00	1.00
176	T2-H2'	T2-H4'	2.44	4.13	M	3.50	2.00
177	C3-H2'	C3-H4'	2.55	4.18	M	3.50	2.00
178	A4-H2'	A4-H4'	2.75	3.64	W	4.50	3.50
179	T5-H2'	T5-H4'	2.18	3.56	W	4.50	3.50
180	C6-H2'	C6-H4'	2.12	3.04	M	3.50	2.00
181	A7-H2'	A7-H4'	2.53	3.62	M	3.50	2.00
182	C8-H2'	C8-H4'	2.02	4.04	M	3.50	2.00
183	G9-H2'	G9-H4'	2.72	4.13	S	2.00	1.00
184	T10-H2'	T10-H4'	2.42	4.13	M	3.50	2.00

185	G11-H2'	G11-H4'	2.84	4.35	M/S	3.00	1.00
186	A12-H2'	A12-H4'	2.55	3.86	M	3.50	2.00
187	T13-H2'	T13-H4'	2.15	3.30	W	4.50	3.50
188	G14-H2'	G14-H4'	2.57	4.05	M	3.50	2.00
189	A15-H2'	A15-H4'	2.55	4.05	S	2.00	1.00
190	G16-H2'	G16-H4'	2.25	3.98	S	2.00	1.00
191	C1-H2''	C1-H4'	2.08	3.98	S	2.00	1.00
192	T2-H2''	T2-H4'	2.15	4.13	S	2.00	1.00
193	C3-H2''	C3-H4'	2.12	4.18	M	3.50	2.00
194	A4-H2''	A4-H4'	2.60	3.64	M	3.50	2.00
195	T5-H2''	T5-H4'	1.76	3.56	M	3.50	2.00
196	C6-H2''	C6-H4'	1.64	3.04	M	3.50	2.00
197	A7-H2''	A7-H4'	2.42	3.62	M	3.50	2.00
198	C8-H2''	C8-H4'	1.97	4.04	M	3.50	2.00
199	G9-H2''	G9-H4'	2.47	4.13	M	3.50	2.00
200	T10-H2''	T10-H4'	2.22	4.13	S	2.00	1.00
201	G11-H2''	G11-H4'	2.60	4.35	M	3.50	2.00
202	A12-H2''	A12-H4'	2.44	3.86	S	2.00	1.00
203	T13-H2''	T13-H4'	2.10	3.30	W	4.50	3.50
204	G14-H2''	G14-H4'	2.45	4.05	M/S	3.00	1.00
205	A15-H2''	A15-H4'	2.44	4.05	M	3.50	2.00
206	G16-H2''	G16-H4'	2.08	3.98	M	3.50	2.00
207	T2-CH3	C1-H6	1.58	7.76	S	2.00	1.00
208	T2-CH3	T2-H6	1.58	7.52	S	2.00	1.00
209	T2-CH3	C1-H5	1.58	5.84	M	3.50	2.00
210	T2-CH3	C1-H1'	1.58	5.80	W	4.50	3.50
211	T2-CH3	C3-H5	1.58	5.65	M	3.50	2.00
212	T2-CH3	T2-H3'	1.58	4.79	VW	5.00	4.00
213	T2-CH3	C1-H3'	1.58	4.55	S	2.00	1.00
214	T2-CH3	T2-H4'	1.58	4.13	W	4.50	3.50
215	T2-CH3	C1-H4'	1.58	3.98	W	4.50	3.50
216	T2-CH3	T2-H2''	1.58	2.15	M	3.50	2.00
217	T2-CH3	C1-H2''	1.58	2.08	M	3.50	2.00
218	T2-CH3	T2-H2'	1.58	2.44	M	3.50	2.00
219	T2-CH3	C1-H2'	1.58	2.44	M	3.50	2.00
220	T10-CH3	G9-H8	1.39	7.83	M	3.50	2.00
221	T10-CH3	T10-H6	1.39	7.30	S	2.00	1.00
222	T10-CH3	G9-H1'	1.39	5.94	M	3.50	2.00
223	T10-CH3	G9-H2'	1.39	2.72	S	2.00	1.00
224	T10-CH3	G9-H2''	1.39	2.47	S	2.00	1.00
225	T10-CH3	T10-H2'	1.39	2.42	M	3.50	2.00
226	T13-CH3	A12-H8	1.26	8.14	S	2.00	1.00
227	T13-CH3	T13-H6	1.26	7.01	S	2.00	1.00
228	T13-CH3	A12-H1'	1.26	6.24	M	3.50	2.00
229	T13-CH3	T13-H1'	1.26	5.40	W	4.50	3.50
230	T13-CH3	A12-H3'	1.26	4.86	M	3.50	2.00
231	T13-CH3	A12-H2'	1.26	2.55	S	2.00	1.00
232	T13-CH3	A12-H2''	1.26	2.44	S	2.00	1.00

233	T13-CH3	T13-H2'	1.26	2.15	M/W	4.00	3.00
234	T13-CH3	T13-H2''	1.26	2.10	M/W	4.00	3.00
235	T5-CH3	A4-H8	1.15	8.32	S	2.00	1.00
236	T5-CH3	T5-H6	1.15	6.89	S	2.00	1.00
237	T5-CH3	A4-H1'	1.15	6.29	M	3.50	2.00
238	T5-CH3	T5-H1'	1.15	5.55	W	4.50	3.50
239	T5-CH3	C6-H5	1.15	5.25	M/W	4.00	3.00
240	T5-CH3	A4-H3'	1.15	4.89	M	3.50	2.00
241	T5-CH3	A4-H2'	1.15	2.75	S	2.00	1.00
242	T5-CH3	A4-H2''	1.15	2.60	S	2.00	1.00
243	T5-CH3	T5-H2'	1.15	2.18	W	4.50	3.50
244	A4-H2	A4-H1'	7.73	6.29	VW	5.00	4.00
245	A4-H2	T13-H1'	7.73	5.40	M	3.50	2.00
246	A7-H2	A7-H1'	7.61	5.88	M	3.50	2.00
247	A7-H2	T10-H1'	7.61	5.44	W	4.50	3.50
248	A12-H2	A12-H1'	8.02	6.24	W	4.50	3.50
249	A12-H2	T5-H1'	8.02	5.55	M	3.50	2.00
250	A12-H2	C6-H1'	8.02	5.61	M	3.50	2.00
251	A12-H2	T13-H1'	8.02	5.40	W	4.50	3.50
252	A15-H2	A15-H1'	7.71	5.77	M/W	4.00	3.00
253	A12-H2	T2-H1'	8.02	5.93	W	4.50	3.50
254	A15-H2	C3-H1'	7.71	5.99	W	4.50	3.50
255	T2-H4'	C1-H1'	4.13	6.29	M	3.50	2.00
256	C3-H4'	T2-H1'	4.18	5.55	M	3.50	2.00
257	A4-H4'	C3-H1'	3.64	5.61	S	2.00	1.00
258	T5-H4'	A4-H1'	3.56	5.88	M	3.50	2.00
259	C6-H4'	T5-H1'	3.04	5.97	M	3.50	2.00
260	A7-H4'	C6-H1'	3.62	5.94	M	3.50	2.00
261	C8-H4'	A7-H1'	4.04	5.44	M	3.50	2.00
262	T10-H4'	G9-H1'	4.13	6.08	S	2.00	1.00
263	G11-H4'	T10-H1'	4.35	6.24	M/W	4.00	3.00
264	A12-H4'	G11-H1'	3.86	5.40	M	3.50	2.00
265	T13-H4'	A12-H1'	3.30	5.94	M	3.50	2.00
266	G14-H4'	T13-H1'	4.05	5.77	M	3.50	2.00
267	A15-H4'	G14-H1'	4.05	5.92	S	2.00	1.00
268	G16-H4'	A15-H1'	3.98	5.77	S	2.00	1.00

Master Table Showing Drug-DNA Connectivities for the 5'*d*-(CTCATCAC).(GTGAGAG)-SJG-136 Adduct

G-5' end

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper Limit	Lower Limit
1	H1a'	A7-H2	3.22	7.61	S	2.00	1.00
2	H1a'	A7-H1'	3.22	5.88	M	3.50	2.00
3	H1a'	A12-H4'	3.22	3.86	M	3.50	2.00
4	H1b'	C8-H1'	2.64	5.97	M	3.50	2.00
5	H1b'	A7-H1'	2.64	5.88	M	3.50	2.00
6	H1b'	A12-H4'	2.64	3.86	M	3.50	2.00
7	H1b'	A7-H2	2.64	7.61	M	3.50	2.00
8	H2a'	C8-H1'	5.24	5.97	W	4.50	3.50
9	H2a'	A7-H2	5.24	7.61	W	4.50	3.50
10	H2a'	G11-H1'	5.24	6.08	S	2.00	1.00
11	H2a'	G11-H4'	5.24	4.35	M	3.50	2.00
12	H2a'	A12-H4'	5.24	3.86	W	4.50	3.50
13	H2b'	A7-H2	5.22	7.61	W	4.50	3.50
14	H2b'	G11-H1'	5.22	6.08	S	2.00	1.00
15	H2b'	G11-H4'	5.22	4.35	M	3.50	2.00
16	H2b'	A12-H4'	5.22	3.86	M	3.50	2.00
17	H3a'	A7-H1'	4.09/4.14	5.88	W	4.50	3.50
18	H3b'	A7-H1'	4.09/4.14	5.88	W	4.50	3.50
19	H6'	A7-H4'	7.06	3.62	M	3.50	2.00
20	H6'	T13-H4'	7.06	3.30	W	4.50	3.50
21	H9'	C6-H1'	6.58	5.61	M	3.50	2.00
22	H9'	A12-H2	6.58	8.02	S	2.00	1.00
23	H9'	A12-H1'	6.58	6.24	M/W	4.00	3.00
24	H9'	T13-H1'	6.58	5.40	M/W	4.00	3.00
25	H9'	T13-H4'	6.58	3.30	VW	5.00	4.00
26	H11'	A7-H2	5.76	7.61	W	4.50	3.50
27	H11'	A7-H4'	5.76	3.62	VW	5.00	4.00
28	H11a'	C6-H1'	3.91	5.61	M/S	3.00	1.00
29	H11a'	A7-H2	3.91	7.61	M	3.50	2.00
30	H11a'	A7-H1'	3.91	5.88	S	2.00	1.00
31	H12a'	C6-H1'	4.20	5.61	M/S	3.00	1.00
32	H12a'	A12-H2	4.20	8.02	M/W	4.00	3.00
33	H12a'	T13-H1'	4.20	5.40	M	3.50	2.00
34	H12b'	A4-H2	4.30	7.73	VW	5.00	4.00
35	H12b'	C6-H1'	4.30	5.61	M/S	3.00	1.00
36	H12b'	A12-H2	4.30	8.02	M/W	4.00	3.00
37	H12b'	T13-H1'	4.30	5.40	M	3.50	2.00
38	H13a/b'	A12-H2	2.30/2.07	8.02	M	3.50	2.00
39	H13a/b'	T13-H4'	2.30/2.07	3.30	W	4.50	3.50
40	H13a/b'	G14-H4'	2.30/2.07	4.05	M/S	3.00	1.00
41	H14'	A7-H3'	3.78	4.78	M	3.50	2.00

42	H14'	A7-H4'	3.78	3.62	M/S	3.00	1.00

G-3' end

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper Limit	Lower Limit
1	H1a	A15-H2	2.95	7.71	W	4.50	3.50
2	H1a	A15-H1'	2.95	7.83	S	2.00	1.00
3	H1a	G16-H4'	2.95	3.98	VW	5.00	4.00
4	H1b	A15-H2	2.58	7.71	M	3.50	2.00
5	H1b	A15-H1'	2.58	7.83	S	2.00	1.00
6	H1b	G16-H4'	2.58	3.98	VW	5.00	4.00
7	H2a	A4-H4'	5.32	3.64	M	3.50	2.00
8	H2a	C3-H1'	5.32	7.45	W	4.50	3.50
9	H2b	A4-H4'	5.17	3.64	M	3.50	2.00
10	H3a	A15-H1'	4.32	7.83	M	3.50	2.00
11	H3b	A4-H1'	4.11	8.32	M	3.50	2.00
12	H6	T5-H4'	6.95	3.56	S	2.00	1.00
13	H9	A4-H2	5.97	7.73	S	2.00	1.00
14	H9	A4-H1'	5.97	8.32	W	4.50	3.50
15	H9	T5-H1'	5.97	6.89	M	3.50	2.00
16	H9	T13-H1'	5.97	7.01	W	4.50	3.50
17	H11	A4-H1'	4.52	8.32	M/S	3.00	1.00
18	H11a	A15-H4'	4.74	4.05	M	3.50	2.00
19	H12a/b	A4-H2	4.25/3.76	7.73	M/S	3.00	1.00
20	H12a/b	T5-H1'	4.25/3.76	6.89	M	3.50	2.00
21	H12a/b	A12-H2	4.25/3.76	8.02	M/W	4.00	3.00
22	H12a/b	T13-H1'	4.25/3.76	7.01	M	3.50	2.00
23	H13a/b	A12-H2	2.07/2.30	8.02	M	3.50	2.00
24	H13a/b	T13-H4'	2.07/2.30	3.30	W	4.50	3.50

Appendix III

Master table showing all restraints for the 5'*d*(CGATTAATCG)$_2$-adozelesin Watson-Crick adduct

Peak number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper limit	Lower limit
1	C1-H6	G2-H8	7.40	7.76	W	4.50	3.50
2	G2-H8	A3-H8	7.76	8.11	W	4.50	3.50
3	A3-H8	T4-H6	8.11	7.05	W	4.50	3.50
4	T4-H6	T5-H6	7.05	7.29	W	4.50	3.50
5	T5-H6	A6-H8	7.29	8.22	W	4.50	3.50
6	A6-H8	A7-H8	8.22	7.81	W	4.50	3.50
7	A7-H8	T8-H6	7.81	6.51	W	4.50	3.50
8	T8-H6	C9-H6	6.51	7.96	M/W	4.00	3.00
9	C9-H6	G10-H8	7.96	7.56	W	4.50	3.50
10	C1-H6	C1-H1'	7.40	5.99	W	4.50	3.50
11	G2-H8	G2-H1'	7.76	5.29	M	3.50	2.00
12	A3-H8	A3-H1'	8.11	6.13	M	3.50	2.00
13	T4-H6	T4-H1'	7.05	5.79	M/W	4.00	3.00
14	T5-H6	T5-H1'	7.29	5.34	M/W	4.00	3.00
15	A6-H8	A6-H1'	8.22	5.90	W	4.50	3.50
16	A7-H8	A7-H1'	7.81	5.78	W	4.50	3.50
17	T8-H6	T8-H1'	6.51	5.18	VW	5.00	4.00
18	C9-H6	C9-H1'	7.96	5.38	W	4.50	3.50
19	G10-H8	G10-H1'	7.56	5.73	M	3.50	2.00
20	G2-H8	C1-H1'	7.76	5.99	M/S	3.00	1.00
21	A3-H8	G2-H1'	8.11	5.29	M	3.50	2.00
22	T4-H6	A3-H1'	7.05	6.13	M	3.50	2.00
23	T5-H6	T4-H1'	7.29	5.79	M/W	4.00	3.00
24	A6-H8	T5-H1'	8.22	5.34	M/W	4.00	3.00
25	A7-H8	A6-H1'	7.81	5.90	M/W	4.00	3.00
26	T8-H6	A7-H1'	6.51	5.78	W	4.50	3.50
27	C9-H6	T8-H1'	7.96	5.18	M	3.50	2.00
28	G10-H8	C9-H1'	7.56	5.38	W	4.50	3.50
29	C1-H6	C1-H3'	7.40	4.51	M	3.50	2.00
30	G2-H8	G2-H3'	7.76	4.14	W	4.50	3.50
31	A3-H8	A3-H3'	8.11	4.89	M/W	4.00	3.00
32	T4-H6	T4-H3'	7.05	4.03	W	4.50	3.50
33	T5-H6	T5-H3'	7.29	4.21	W	4.50	3.50
34	A6-H8	A6-H3'	8.22	4.85	W	4.50	3.50
35	A7-H8	A7-H3'	7.81	4.40	M/W	4.00	3.00
36	T8-H6	T8-H3'	6.51	4.21	W	4.50	3.50
37	C9-H6	C9-H3'	7.96	4.29	W	4.50	3.50
38	G10-H8	G10-H3'	7.56	4.34	M	3.50	2.00
39	G2-H8	C1-H3'	7.76	4.51	M	3.50	2.00

40	A3-H8	G2-H3'	8.11	4.14	W	4.50	3.50
41	T4-H6	A3-H3'	7.05	4.89	W	4.50	3.50
42	T5-H6	T4-H3'	7.29	4.03	M	3.50	2.00
43	A6-H8	T5-H3'	8.22	4.21	W	4.50	3.50
44	A7-H8	A6-H3'	7.81	4.85	W	4.50	3.50
45	T8-H6	A7-H3'	6.51	4.40	VW	5.00	4.00
46	C9-H6	T8-H3'	7.96	4.21	W	4.50	3.50
47	G10-H8	C9-H3'	7.56	4.29	W	4.50	3.50
48	C1-H6	G2-H4'	7.40	3.04	M	3.50	2.00
49	G2-H8	A3-H4'	7.76	3.90	M/W	4.00	3.00
50	A3-H8	A3-H4'	8.11	3.90	M/W	4.00	3.00
51	T4-H6	T4-H4'	7.05	3.49	M	3.50	2.00
52	T5-H6	T5-H4'	7.29	3.31	W	4.50	3.50
53	A6-H8	A6-H4'	8.22	3.76	W	4.50	3.50
54	A7-H8	A7-H4'	7.81	3.00	M/W	4.00	3.00
55	T8-H6	T8-H4'	6.51	4.00	VW	5.00	4.00
56	C9-H6	C9-H4'	7.96	3.80	M/W	4.00	3.00
57	G10-H8	G10-H4'	7.56	3.51	W	4.50	3.50
58	C1-H6	C1-H2'	7.40	2.19	M	3.50	2.00
59	G2-H8	G2-H2'	7.76	2.55	S	2.00	1.00
60	A3-H8	A3-H2'	8.11	2.57	S	2.00	1.00
61	T4-H6	T4-H2'	7.05	1.77	M/S	3.00	1.00
62	T5-H6	T5-H2'	7.29	2.79	M/S	3.00	1.00
63	A6-H8	A6-H2'	8.22	2.53	M/S	3.00	1.00
64	A7-H8	A7-H2'	7.81	2.79	M/S	3.00	1.00
65	T8-H6	T8-H2'	6.51	1.34	W	4.50	3.50
66	C9-H6	C9-H2'	7.96	1.49	M	3.50	2.00
67	G10-H8	G10-H2'	7.56	2.07	S	2.00	1.00
68	C1-H6	C1-H2''	7.40	2.42	S	2.00	1.00
69	G2-H8	G2-H2''	7.76	2.64	S	2.00	1.00
70	A3-H8	A3-H2''	8.11	2.78	S	2.00	1.00
71	T4-H6	T4-H2''	7.05	2.34	S	2.00	1.00
72	T5-H6	T5-H2''	7.29	2.41	M	3.50	2.00
73	A6-H8	A6-H2''	8.22	2.66	M	3.50	2.00
74	A7-H8	A7-H2''	7.81	2.41	M/S	3.00	1.00
75	T8-H6	T8-H2''	6.51	1.84	M/W	4.00	3.00
76	C9-H6	C9-H2''	7.96	1.89	M	3.50	2.00
77	G10-H8	G10-H2''	7.56	2.31	M/S	3.00	1.00
78	G2-H8	C1-H2'	7.76	2.19	S	2.00	1.00
79	A3-H8	G2-H2'	8.11	2.55	S	2.00	1.00
80	T4-H6	A3-H2'	7.05	2.57	M/S	3.00	1.00
81	T5-H6	T4-H2'	7.29	1.77	M	3.50	2.00
82	A6-H8	T5-H2'	8.22	2.79	M	3.50	2.00
83	A7-H8	A6-H2'	7.81	2.53	S	2.00	1.00
84	T8-H6	A7-H2'	6.51	2.79	M/W	4.00	3.00
85	C9-H6	T8-H2'	7.96	1.34	M/S	3.00	1.00
86	G10-H8	C9-H2'	7.56	1.49	M	3.50	2.00
87	G2-H8	C1-H2''	7.76	2.42	S	2.00	1.00

88	A3-H8	G2-H2''	8.11	2.64	S	2.00	1.00
89	T4-H6	A3-H2''	7.05	2.78	M/S	3.00	1.00
90	T5-H6	T4-H2''	7.29	2.34	M	3.50	2.00
91	A6-H8	T5-H2''	8.22	2.41	M	3.50	2.00
92	A7-H8	A6-H2''	7.81	2.66	S	2.00	1.00
93	T8-H6	A7-H2''	6.51	2.41	W	4.50	3.50
94	C9-H6	T8-H2''	7.96	1.84	M/S	3.00	1.00
95	G10-H8	C9-H2''	7.56	1.89	M/W	4.00	3.00
96	C1-H1'	C1-H3'	7.76	7.40	M	3.50	2.00
97	G2-H1'	G2-H3'	8.11	7.76	M	3.50	2.00
98	A3-H1'	A3-H3'	7.05	8.11	W	4.50	3.50
99	T4-H1'	T4-H3'	7.29	7.05	M/W	4.00	3.00
100	T5-H1'	T5-H3'	8.22	7.29	M	3.50	2.00
101	A6-H1'	A6-H3'	7.81	8.22	M	3.50	2.00
102	A7-H1'	A7-H3'	6.51	7.81	M/W	4.00	3.00
103	T8-H1'	T8-H3'	7.96	6.51	W	4.50	3.50
104	C9-H1'	C9-H3'	7.56	7.96	W	4.50	3.50
105	G10-H1'	G10-H3'	5.73	7.56	M	3.50	2.00
106	C1-H1'	C1-H4'	7.76	3.80	M	3.50	2.00
107	G2-H1'	G2-H4'	8.11	3.04	M	3.50	2.00
108	A3-H1'	A3-H4'	7.05	3.90	M	3.50	2.00
109	T4-H1'	T4-H4'	7.29	3.90	M/W	4.00	3.00
110	T5-H1'	T5-H4'	8.22	3.49	M	3.50	2.00
111	A6-H1'	A6-H4'	7.81	3.31	M	3.50	2.00
112	A7-H1'	A7-H4'	6.51	3.76	M	3.50	2.00
113	C9-H1'	C9-H4'	7.56	4.00	M	3.50	2.00
114	G10-H1'	G10-H4'	5.73	3.51	M/S	3.00	1.00
115	C1-H3'	C1-H4'	7.40	3.80	M	3.50	2.00
116	G2-H3'	G2-H4'	7.76	3.04	M	3.50	2.00
117	A3-H3'	A3-H4'	8.11	3.90	M	3.50	2.00
118	T4-H3'	T4-H4'	7.05	3.90	M	3.50	2.00
119	T5-H3'	T5-H4'	7.29	3.49	M	3.50	2.00
120	A6-H3'	A6-H4'	8.22	3.31	M	3.50	2.00
121	A7-H3'	A7-H4'	7.81	3.76	M	3.50	2.00
122	T8-H3'	T8-H4'	6.51	3.00	M	3.50	2.00
123	C9-H3'	C9-H4'	7.96	4.00	W	4.50	3.50
124	G10-H3'	G10-H4'	7.56	3.51	M/S	3.00	1.00
125	C1-H2'	C1-H3'	2.19	7.40	W	4.50	3.50
126	G2-H2'	G2-H3'	2.55	7.76	M	3.50	2.00
127	A3-H2'	A3-H3'	2.57	8.11	M	3.50	2.00
128	T4-H2'	T4-H3'	1.77	7.05	M	3.50	2.00
129	A6-H2'	A6-H3'	2.53	8.22	M	3.50	2.00
130	A7-H2'	A7-H3'	2.79	7.81	W	4.50	3.50
131	T8-H2'	T8-H3'	1.34	6.51	M	3.50	2.00
132	C9-H2'	C9-H3'	1.49	7.96	M	3.50	2.00
133	G10-H2'	G10-H3'	2.07	7.56	M	3.50	2.00
134	C1-H2''	C1-H3'	2.42	7.40	M	3.50	2.00
135	G2-H2''	G2-H3'	2.64	7.76	M	3.50	2.00

136	A3-H2''	A3-H3'	2.78	8.11	M	3.50	2.00
137	T4-H2''	T4-H3'	2.34	7.05	M	3.50	2.00
138	T5-H2''	T5-H3'	2.41	7.29	W	450	3.50
139	A6-H2''	A6-H3'	2.66	8.22	M	3.50	2.00
	A7-H2''	A7-H3'	2.41	7.81	W	4.50	3.50
141	T8-H2''	T8-H3'	1.84	6.51	M	3.50	2.00
142	C9-H2''	C9-H3'	1.89	7.96	M	3.50	2.00
143	G10-H2''	G10-H3'	2.31	7.56	M	3.50	2.00
144	C1-H2'	C1-H4'	2.19	3.80	M	3.50	2.00
145	G2-H2'	G2-H4'	2.55	3.04	M	3.50	2.00
146	A3-H2'	A3-H4'	2.57	3.90	M/W	4.00	3.00
147	T4-H2'	T4-H4'	1.77	3.90	W	4.50	3.50
148	T5-H2'	T5-H4'	2.79	3.49	M/W	4.00	3.00
149	A6-H2'	A6-H4'	2.53	3.31	M/W	4.00	3.00
150	A7-H2'	A7-H4'	2.79	3.76	M/W	4.00	3.00
151	T8-H2'	T8-H4'	1.34	3.00	W	4.50	3.50
152	C9-H2'	C9-H4'	1.49	4.00	M/W	4.00	3.00
153	G10-H2'	G10-H4'	2.07	3.51	M	3.50	2.00
154	C1-H2''	C1-H4'	2.42	3.80	M	3.50	2.00
155	G2-H2''	G2-H4'	2.64	3.04	M	3.50	2.00
156	A3-H2''	A3-H4'	2.78	3.90	M/W	4.00	3.00
157	T4-H2''	T4-H4'	2.34	3.90	W	4.50	3.50
158	T5-H2''	T5-H4'	2.41	3.49	M/W	4.00	3.00
159	A6-H2''	A6-H4'	2.66	3.31	M/W	4.00	3.00
160	A7-H2''	A7-H4'	2.41	3.76	M/W	4.00	3.00
161	T8-H2''	T8-H4'	1.84	3.00	W	4.50	3.50
162	C9-H2''	C9-H4'	1.89	4.00	M/W	4.00	3.00
163	G10-H2''	G10-H4'	2.31	3.51	M	3.50	2.00
164	T8-CH3	A7-H3'	0.97	4.40	W	4.50	3.50
165	T8-CH3	T8-H1'	0.97	5.18	VW	5.00	4.00
166	T8-CH3	T8-H6	0.97	6.51	M/S	3.00	1.00
167	T8-CH3	A7-H8	0.97	7.81	M	3.50	2.00
168	T8-CH3	T8-H3'	0.97	4.21	VW	5.00	4.00
169	T8-CH3	A7-H2'	0.97	2.09	M	3.50	2.00
170	T8-CH3	A7-H2''	0.97	2.41	M	3.50	2.00
171	T8-CH3	A7-H1'	0.97	5.78	W	4.50	3.50
172	T4-CH3	T4-H6	1.11	7.05	M/S	3.00	1.00
173	T4-CH3	A3-H8	1.11	8.11	S	2.00	1.00
174	T4-CH3	T4-H1'	1.11	5.79	W	4.50	3.50
175	T4-CH3	A3-H1'	1.11	6.13	M	3.50	2.00
176	T4-CH3	A3-H3'	1.11	4.89	M	3.50	2.00
177	T4-CH3	A3-H2'	1.11	2.57	S	2.00	1.00
178	T4-CH3	A3-H2''	1.11	2.78	S	2.00	1.00
179	T4-CH3	T4-H3'	1.11	4.03	W	4.50	3.50
180	T4-CH3	T4-H2'	1.11	1.77	W	4.50	3.50
181	T4-CH3	T4-H2''	1.11	2.34	M	3.50	2.00
182	T5-CH3	T5-H6	1.49	7.29	M/S	3.00	1.00
183	T5-CH3	T4-H6	1.49	7.05	M/S	3.00	1.00

184	T5-CH3	T4-H1'	1.49	5.79	M	3.50	2.00
185	T5-CH3	T4-H3'	1.49	4.03	VW	5.00	4.00
186	T5-CH3	T4-H2'	1.49	1.77	M/S	3.00	1.00
187	T5-CH3	T4-H2''	1.49	2.34	M	3.50	2.00
188	C1-H5	C1-H6	5.67	7.40	S	2.00	1.00
189	C9-H5	C9-H6	5.18	7.96	S	2.00	1.00
190	A3-H2	A3-H1'	7.40	6.13	VW	5.00	4.00
191	A7-H2	A7-H1'	7.82	5.78	W	4.50	3.50
192	A7-H2	T4-H1'	7.82	5.79	W	4.50	3.50
193	A7-H2	T5-H1'	7.82	5.34	W	4.50	3.50
194	A7-H2	T8-H1'	7.82	5.18	VW	5.00	4.00

Drug-DNA restraints

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Peak intensity	Upper limit	Lower limit
1	Ado-H6	Ado-CH3	8.28	2.70	M	3.50	2.00
2	Ado-CH3	Ado-H8A/B	2.70	5.50	M	3.50	2.00
3	Ado-H8A/B	Ado-H1A/B	5.50	3.51	W	4.50	3.50
4	Ado-H8A/B	Ado-H8A2	5.50	4.20	S	2.00	1.00
5	Ado-H8A2	T4-H1'	4.20	5.79	M/W	4.00	3.00
6	Ado-H8A2	A3-H2	4.20	7.40	M	3.50	2.00
7	Ado-H8A2	A7-H2	4.20	7.82	M	3.50	2.00
8	Ado-H1A/B	A7-H2	3.51	7.82	M	3.50	2.00
9	Ado-H1A/B	A3-H2	3.51	7.40	S	2.00	1.00
10	Ado-H5'2	A7-H1'	7.46	5.78	M	3.50	2.00
11	Ado-H5'2	A6-H8	7.46	8.22	W	4.50	3.50
12	Ado-H5'2	T8-H5'	7.46		W	4.50	3.50
13	Ado-H5'2	T8-H5''	7.46		W	4.50	3.50
14	Ado-H6'2	A7-H1'	7.53	5.78	W	4.50	3.50
15	Ado-H6'2	A7-H2	7.53	7.82	M	3.50	2.00
16	Ado-H6'2	T8-H4'	7.53	4.00	W	4.50	3.50
17	Ado-H6'2	T8-H5'	7.53		W	4.50	3.50
18	Ado-H3'1	T4-H3	7.92	12.80	M/W	4.00	3.00
19	Ado-H3'1	T8-H3	7.92	11.02	M/W	4.00	3.00
20	Ado-H3'1	A3-H2	7.92	7.40	M	3.50	2.00
21	Ado-H3'1	T4-H1'	7.92	5.79	W	4.50	3.50
22	Ado-H3'1	T5-H1'	7.92	5.34	VW	5.00	4.00
23	Ado-H3	T4-H1'	8.14	5.79	VW	5.00	4.00
24	Ado-H4'1	T5-H3	6.96	9.76	M/W	4.00	3.00
25	Ado-H4'2	T5-H3	7.80	9.76	M/W	4.00	3.00
26	Ado-H4'2	A6-H1'	7.80	5.90	VW	5.00	4.00
27	Ado-H4'2	T8-H5'	7.80		VW	5.00	4.00
28	Ado-H4'2	T8-H5''	7.80		VW	5.00	4.00
29	Ado-H7'2	T8-H5'	8.10		M	3.50	2.00

30	Ado-H7'2	T8-H5''	8.10		VW	5.00	4.00

Appendix IV

Master table showing all restraints for the 5'*d*(CGATTAATCG)$_2$-adozelesin Hoogsteen adduct

Peak number	Proton 1	Proton 2	Shift 1	Shift 2	Intensity	Upper limit	Lower limit
1	C1-H6	G2-H8	7.41	7.79	W	4.50	3.50
2	G2-H8	A3-H8	7.79	8.15	W	4.50	3.50
3	A3-H8	T4-H6	8.15	6.90	W	4.50	3.50
4	T4-H6	T5-H6	6.90	6.92	W	4.50	3.50
5	T5-H6	A6-H8	6.92	7.80	W	4.50	3.50
6	A6-H8	A7-H8	7.80	8.14	W	4.50	3.50
7	A7-H8	T8-H6	8.14	6.90	W	4.50	3.50
8	T8-H6	C9-H6	6.90	7.29	W	4.50	3.50
9	C9-H6	G10-H8	7.29	7.41	W	4.50	3.50
10	C1-H6	C1-H1'	7.41	5.51	M	3.50	2.00
11	G2-H8	G2-H1'	7.79	5.34	M/W	4.00	3.00
12	A3-H8	A3-H1'	8.15	5.63	W	4.50	3.50
13	T4-H6	T4-H1'	6.90	5.58	M/W	4.00	3.00
14	T5-H6	T5-H1'	6.92	5.58	M/W	4.00	3.00
15	A6-H8	A6-H1'	7.80	5.39	M	3.50	2.00
16	A7-H8	A7-H1'	8.14	6.23	M/W	4.00	3.00
17	T8-H6	T8-H1'	6.90	5.74	M	3.50	2.00
18	C9-H6	C9-H1'	7.29	5.51	M	3.50	2.00
19	G10-H8	G10-H1'	7.41	5.68	M	3.50	2.00
20	G2-H8	C1-H1'	7.79	5.51	M	3.50	2.00
21	A3-H8	G2-H1'	8.15	5.34	W	4.50	3.50
22	T4-H6	A3-H1'	6.90	5.63	M/W	4.00	3.00
23	T5-H6	T4-H1'	6.92	5.58	M	3.50	2.00
24	A6-H8	T5-H1'	7.80	5.58	M	3.50	2.00
25	A7-H8	A6-H1'	8.14	5.39	M	3.50	2.00
26	T8-H6	A7-H1'	6.90	6.23	M	3.50	2.00
27	C9-H6	T8-H1'	7.29	5.74	M	3.50	2.00
28	G10-H8	C9-H1'	7.41	5.51	M	3.50	2.00
29	C1-H6	C1-H3'	7.41	4.69	W	4.50	3.50
30	G2-H8	G2-H3'	7.79	4.23	W	4.50	3.50
31	A3-H8	A3-H3'	8.15	4.97	W	4.50	3.50
32	T4-H6	T4-H3'	6.90	4.91	W	4.50	3.50
33	T5-H6	T5-H3'	6.92	4.28	M	3.50	2.00
34	A6-H8	A6-H3'	7.80	4.88	W	4.50	3.50
35	A7-H8	A7-H3'	8.14	4.40	W	4.50	3.50
36	T8-H6	T8-H3'	6.90	4.09	W	4.50	3.50
37	C9-H6	C9-H3'	7.29	3.96	M	3.50	2.00
38	G10-H8	G10-H3'	7.41	4.40	W	4.50	3.50
39	G2-H8	C1-H3'	7.79	4.69	W	4.50	3.50

40	A3-H8	G2-H3'	8.15	4.23	VW	5.00	4.00
41	T4-H6	A3-H3'	6.90	4.97	W	4.50	3.50
42	T5-H6	T4-H3'	6.92	4.91	W	4.50	3.50
43	A6-H8	T5-H3'	7.80	4.28	W	4.50	3.50
44	A7-H8	A6-H3'	8.14	4.88	W	4.50	3.50
45	C9-H6	T8-H3'	7.29	4.09	M	3.50	2.00
46	G10-H8	C9-H3'	7.41	3.96	W	4.50	3.50
47	C1-H6	C1-H4'	7.41	3.89	M	3.50	2.00
48	A3-H8	A3-H4'	8.15	3.91	W	4.50	3.50
49	T4-H6	T4-H4'	6.90	3.50	W	4.50	3.50
50	T5-H6	T5-H4'	6.92	3.32	VW	5.00	4.00
51	T8-H6	T8-H4'	6.90	4.01	W	4.50	3.50
52	C9-H6	C9-H4'	7.29	3.89	M/W	4.00	3.00
53	G10-H8	G10-H4'	7.41	3.55	W	4.50	3.50
54	C1-H6	C1-H2'	7.41	1.63	S	2.00	1.00
55	G2-H8	G2-H2'	7.79	2.60	S	2.00	1.00
56	A3-H8	A3-H2'	8.15	2.24	S	2.00	1.00
57	T4-H6	T4-H2'	6.90	1.25	M	3.50	2.00
58	T5-H6	T5-H2'	6.92	1.62	M	3.50	2.00
59	A6-H8	A6-H2'	7.80	2.57	S	2.00	1.00
60	A7-H8	A7-H2'	8.14	2.51	S	2.00	1.00
61	T8-H6	T8-H2'	6.90	1.82	S	2.00	1.00
62	C9-H6	C9-H2'	7.29	1.84	S	2.00	1.00
63	G10-H8	G10-H2'	7.41	1.68	S	2.00	1.00
64	C1-H6	C1-H2''	7.41	2.22	S	2.00	1.00
65	G2-H8	G2-H2''	7.79	2.70	S	2.00	1.00
66	A3-H8	A3-H2''	8.15	2.62	M	3.50	2.00
67	T4-H6	T4-H2''	6.90	1.40	M	3.50	2.00
68	T5-H6	T5-H2''	6.92	2.08	S	2.00	1.00
69	A6-H8	A6-H2''	7.80	2.70	S	2.00	1.00
70	A7-H8	A7-H2''	8.14	2.59	S	2.00	1.00
71	T8-H6	T8-H2''	6.90	2.12	S	2.00	1.00
72	C9-H6	C9-H2''	7.29	2.22	S	2.00	1.00
73	G10-H8	G10-H2''	7.41	2.09	S	2.00	1.00
74	G2-H8	C1-H2'	7.79	1.63	S	2.00	1.00
75	A3-H8	G2-H2'	8.15	2.60	S	2.00	1.00
76	T4-H6	A3-H2'	6.90	2.24	M	3.50	2.00
77	T5-H6	T4-H2'	6.92	1.25	M	3.50	2.00
78	A6-H8	T5-H2'	7.80	1.62	M	3.50	2.00
79	A7-H8	A6-H2'	8.14	2.57	M/S	3.00	1.00
80	T8-H6	A7-H2'	6.90	2.51	M	3.50	2.00
81	C9-H6	T8-H2'	7.29	1.82	M	3.50	2.00
82	G10-H8	C9-H2'	7.41	1.84	M/S	3.00	1.00
83	G2-H8	C1-H2''	7.79	2.22	M	3.50	2.00
84	A3-H8	G2-H2''	8.15	2.70	S	2.00	1.00
85	T4-H6	A3-H2''	6.90	2.62	S	2.00	1.00
86	T5-H6	T4-H2''	6.92	1.40	M	3.50	2.00
87	A6-H8	T5-H2''	7.80	2.08	M	3.50	2.00

88	A7-H8	A6-H2''	8.14	2.70	S	2.00	1.00
89	T8-H6	A7-H2''	6.90	2.59	M	3.50	2.00
90	C9-H6	T8-H2''	7.29	2.12	S	2.00	1.00
91	G10-H8	C9-H2''	7.41	2.22	M/S	3.00	1.00
92	C1-H1'	C1-H3'	5.51	4.69	W	4.50	3.50
93	G2-H1'	G2-H3'	5.34	4.23	M	3.50	2.00
94	A3-H1'	A3-H3'	5.63	4.97	W	4.50	3.50
95	T5-H1'	T5-H3'	5.58	4.28	M/W	4.00	3.00
96	A6-H1'	A6-H3'	5.39	4.88	W	4.50	3.50
97	A7-H1'	A7-H3'	6.23	4.40	W	4.50	3.50
98	T8-H1'	T8-H3'	5.74	4.09	M	3.50	2.00
99	C9-H1'	C9-H3'	5.51	3.96	W	4.50	3.50
100	G10-H1'	G10-H3'	5.68	4.40	W	4.50	3.50
101	C1-H1'	C1-H4'	5.51	3.89	M	3.50	2.00
102	A3-H1'	A3-H4'	5.63	3.91	M	3.50	2.00
103	T4-H1'	T4-H4'	5.58	3.50	W	4.50	3.50
104	A6-H1'	A6-H4'	5.39	3.80	M	3.50	2.00
105	A7-H1'	A7-H4'	6.23	3.32	W	4.50	3.50
106	T8-H1'	T8-H4'	5.74	4.01	M	3.50	2.00
107	C9-H1'	C9-H4'	5.51	3.89	M	3.50	2.00
108	G10-H1'	G10-H4'	5.68	3.55	W	4.50	3.50
109	A3-H3'	A3-H4'	4.97	3.91	W	4.50	3.50
110	T4-H3'	T4-H4'	4.91	3.50	W	4.50	3.50
111	T5-H3'	T5-H4'	4.28	3.32	M	3.50	2.00
112	A6-H3'	A6-H4'	4.88	3.80	M/W	4.00	3.00
113	A7-H3'	A7-H4'	4.40	3.32	M/W	4.00	3.00
114	T8-H3'	T8-H4'	4.09	4.01	M	3.50	2.00
115	C9-H3'	C9-H4'	3.96	3.89	M	3.50	2.00
116	C1-H2'	C1-H3'	1.63	4.69	W	4.50	3.50
117	G2-H2'	G2-H3'	2.60	4.23	M	3.50	2.00
118	A3-H2'	A3-H3'	2.24	4.97	M	3.50	2.00
119	T4-H2'	T4-H3'	1.25	4.91	M	3.50	2.00
120	T5-H2'	T5-H3'	1.62	4.28	M	3.50	2.00
121	A6-H2'	A6-H3'	2.57	4.88	M	3.50	2.00
122	A7-H2'	A7-H3'	2.51	4.40	W	4.50	3.50
123	T8-H2'	T8-H3'	1.82	4.09	W	4.50	3.50
124	C9-H2'	C9-H3'	1.84	3.96	M	3.50	2.00
125	G10-H2'	G10-H3'	1.68	4.40	M	3.50	2.00
126	C1-H2''	C1-H3'	2.22	4.69	VW	5.00	4.00
127	G2-H2''	G2-H3'	2.70	4.23	M	3.50	2.00
128	A3-H2''	A3-H3'	2.62	4.97	M	3.50	2.00
129	T4-H2''	T4-H3'	1.40	4.91	M	3.50	2.00
130	T5-H2''	T5-H3'	2.08	4.28	M	3.50	2.00
131	A6-H2''	A6-H3'	2.70	4.88	M	3.50	2.00
132	A7-H2''	A7-H3'	2.59	4.40	M	3.50	2.00
133	T8-H2''	T8-H3'	2.12	4.09	M	3.50	2.00
134	C9-H2''	C9-H3'	2.22	3.96	M	3.50	2.00
135	G10-H2''	G10-H3'	2.09	4.40	M	3.50	2.00

136	C1-H2'	C1-H4'	1.63	3.89	M	3.50	2.00
137	A3-H2'	A3-H4'	2.24	3.91	M	3.50	2.00
138	T4-H2'	T4-H4'	1.25	3.50	W	4.50	3.50
139	T5-H2'	T5-H4'	1.62	3.32	M	3.50	2.00
140	A6-H2'	A6-H4'	2.57	3.80	M	3.50	2.00
141	A7-H2'	A7-H4'	2.51	3.32	W	4.50	3.50
142	T8-H2'	T8-H4'	1.82	4.01	M	3.50	2.00
143	C9-H2'	C9-H4'	1.84	3.89	M	3.50	2.00
144	G10-H2'	G10-H4'	1.68	3.55	M	3.50	2.00
145	C1-H2''	C1-H4'	2.22	3.89	M	3.50	2.00
146	A3-H2''	A3-H4'	2.62	3.91	M	3.50	2.00
147	T4-H2''	T4-H4'	1.40	3.50	W	4.50	3.50
148	T5-H2''	T5-H4'	2.08	3.32	W	4.50	3.50
149	A6-H2''	A6-H4'	2.70	3.80	M	3.50	2.00
150	A7-H2''	A7-H4'	2.59	3.32	W	4.50	3.50
151	T8-H2''	T8-H4'	2.12	4.01	M	3.50	2.00
152	C9-H2''	C9-H4'	2.22	3.89	M	3.50	2.00
153	G10-H2''	G10-H4'	2.09	3.55	M	3.50	2.00
154	T8-CH3	A7-H3'	1.13	4.40	W	4.50	3.50
155	T8-CH3	T8-H6	1.13	6.90	S	2.00	1.00
156	T8-CH3	A7-H8	1.13	8.14	M/S	3.00	1.00
157	T8-CH3	T8-H2'	1.13	1.82	W	4.50	3.50
158	T8-CH3	T8-H2''	1.13	2.12	M	3.50	2.00
159	T8-CH3	A7-H2'	1.13	2.51	S	2.00	1.00
160	T8-CH3	A7-H2''	1.13	2.59	S	2.00	1.00
161	T8-CH3	A7-H1'	1.13	6.23	M	3.50	2.00
162	T4-CH3	T4-H6	1.16	6.90	S	2.00	1.00
163	T4-CH3	A3-H8	1.16	8.15	M/S	3.00	1.00
164	T4-CH3	T4-H1'	1.16	5.58	W	4.50	3.50
165	T4-CH3	A3-H1'	1.16	5.63	M	3.50	2.00
166	T4-CH3	A3-H3'	1.16	4.97	W	4.50	3.50
167	T4-CH3	A3-H2'	1.16	2.24	M/S	3.00	1.00
168	T4-CH3	A3-H2''	1.16	2.62	M/S	3.00	1.00
169	T4-CH3	T4-H3'	1.16	4.91	W	4.50	3.50
170	T5-CH3	T5-H6	1.32	6.92	S	2.00	1.00
171	T5-CH3	T4-H6	1.32	6.90	S	2.00	1.00
172	T5-CH3	T4-H1'	1.32	5.58	M	3.50	2.00
173	T5-CH3	T4-H2'	1.32	1.25	M	3.50	2.00
174	T5-CH3	T4-H2''	1.32	1.40	M	3.50	2.00
175	C1-H5	C1-H6	5.68	7.41	S	2.00	1.00
176	C9-H5	C9-H6	5.45	7.29	S	2.00	1.00
177	C9-H5	T8-CH3	5.45	1.13	M	3.50	2.00
178	A7-H2	A7-H1'	7.82	6.23	W	4.50	3.50
179	A7-H2	T5-H1'	7.82	5.58	W	4.50	3.50
180	A3-H2	T8-H3	7.28	10.13	W	4.50	3.50
181	A3-H2	T4-H3	7.28	11.17	W	4.50	3.50
182	A6-H2	A7-H2	7.43	7.82	M	3.50	2.00
183	A6-H2	T5-H2'	7.43	1.62	M	3.50	2.00

184	A6-H2	T5-H2''	7.43	2.08	M	3.50	2.00

Drug-DNA restraints

Peak Number	Proton 1	Proton 2	Shift 1	Shift 2	Peak intensity	Upper Limit	Lower Limit
1	Ado-H6	Ado-CH3	8.28	2.83	M/S	3.00	1.00
2	Ado-H6	A3-H4'	8.28	3.91	W	4.50	3.50
3	Ado-CH3	A3-H8	2.83	8.15	VW	5.00	4.00
4	Ado-CH3	Ado-H8A/B	2.83	5.35	M	3.50	2.00
5	Ado-H8A/B	Ado-H8A2	5.35	3.90	M	3.50	2.00
6	Ado-H8A/B	Ado-H1A/B	5.35	3.89	M	3.50	2.00
7	Ado-H8A2	A3-H1'	3.90	5.63	W	4.50	3.50
8	Ado-H8A2	A3-H2	3.90	7.28	M	3.50	2.00
9	Ado-H1A/B	A3-H2	3.89	7.28	M	3.50	2.00
10	Ado-H1A/B	T8-H1'	3.89	5.74	VW	5.00	4.00
11	Ado-H6'2	Ado-H4'1	7.38	7.80	M	3.50	2.00
12	Ado-H5'2	T5-H3	8.16	12.90	W	4.50	3.50